Quadratic Estimators of the Power Spectrum

C. T. Mullis and Louis L. Scharf

Department of Electrical and Computer Engineering

UNIVERSITY OF COLORADO

BOULDER, COLORADO

89 5 19 145

(4)

Quadratic Estimators of the Power Spectrum

C. T. Mullis and Louis L. Scharf

TECHNICAL REPORT

APRIL 1989

Quadratic Estimators of the Power Spectrum*

C. T. Mullis and Louis L. Scharf

Electrical and Computer Engineering

University of Colorado

Boulder, CO 80309

* This work supported by the Office of Naval Research under contract No. N00014-89-J-1070, by the Army Research Office under contract No. DAAL03-88-K-0058, and by the National Science Foundation under contract No. MIP-8717714.

Figures

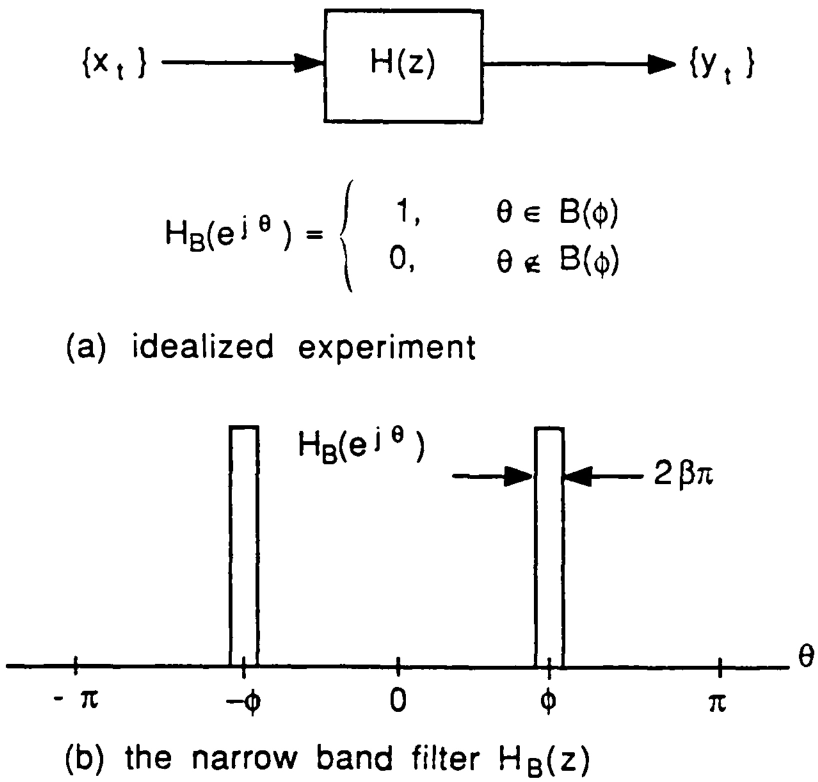
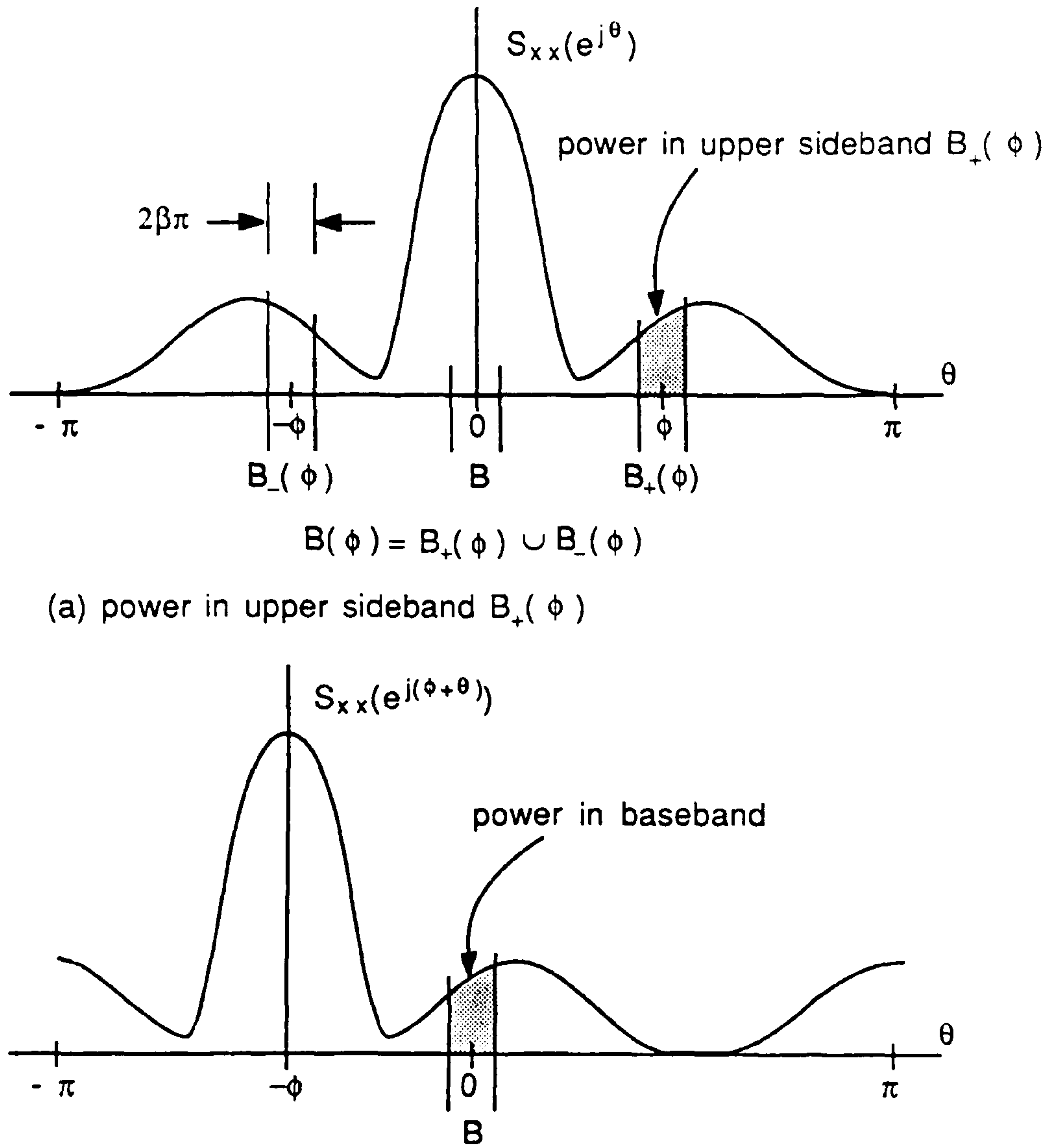


Figure 1: *Idealized Experiment for Spectrum Analysis*



(a) power in upper sideband $B_+(\phi)$

(b) power of demodulated signal in baseband B

Figure 2: *Power in Upper Sideband and in Baseband*

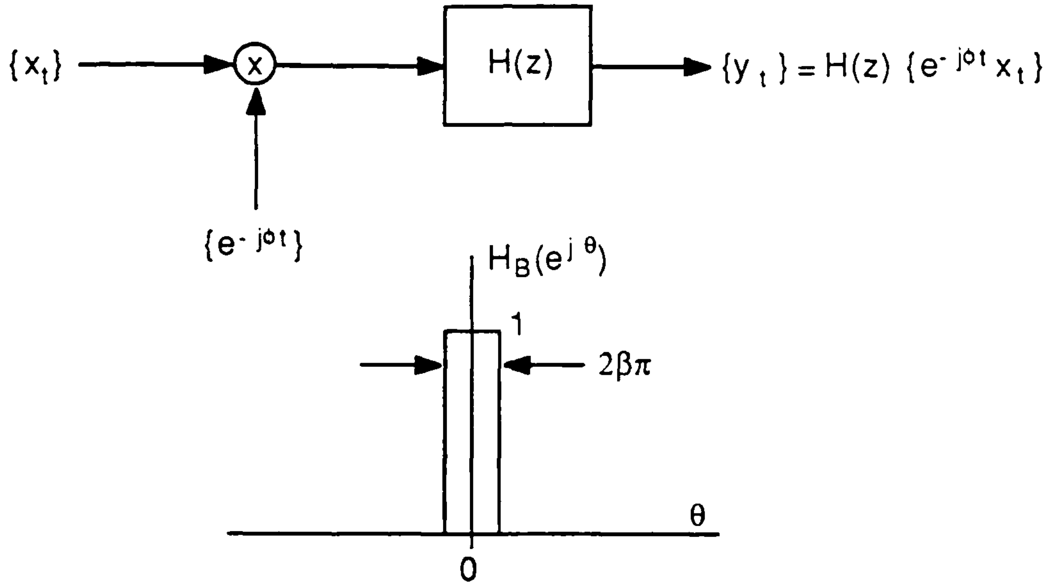
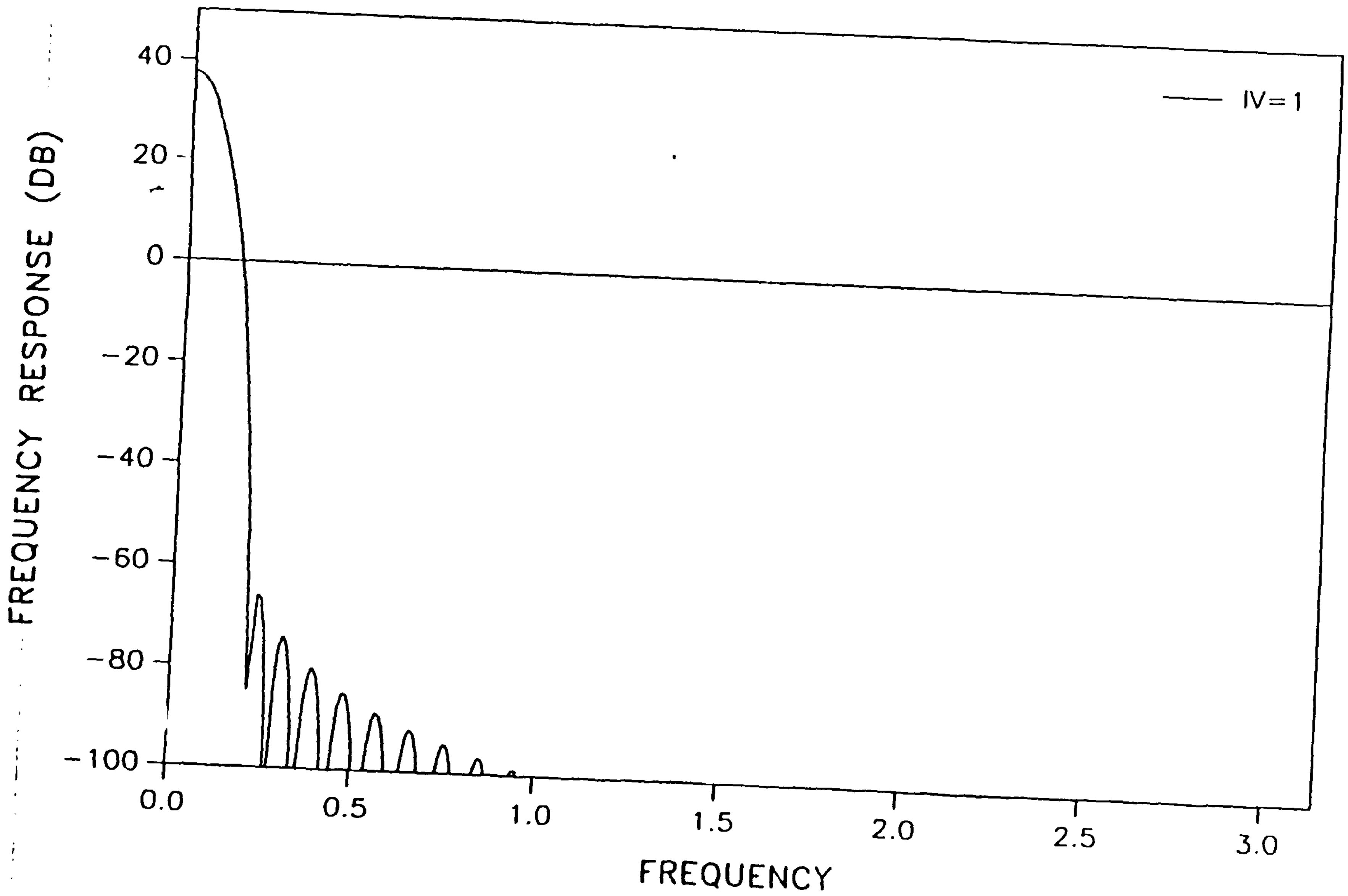


Figure 3: *Alternative Experiment for Spectrum Analysis*

Figure 4: *Frequency Response of Several Dominant and Subdominant Slepian Sequences*

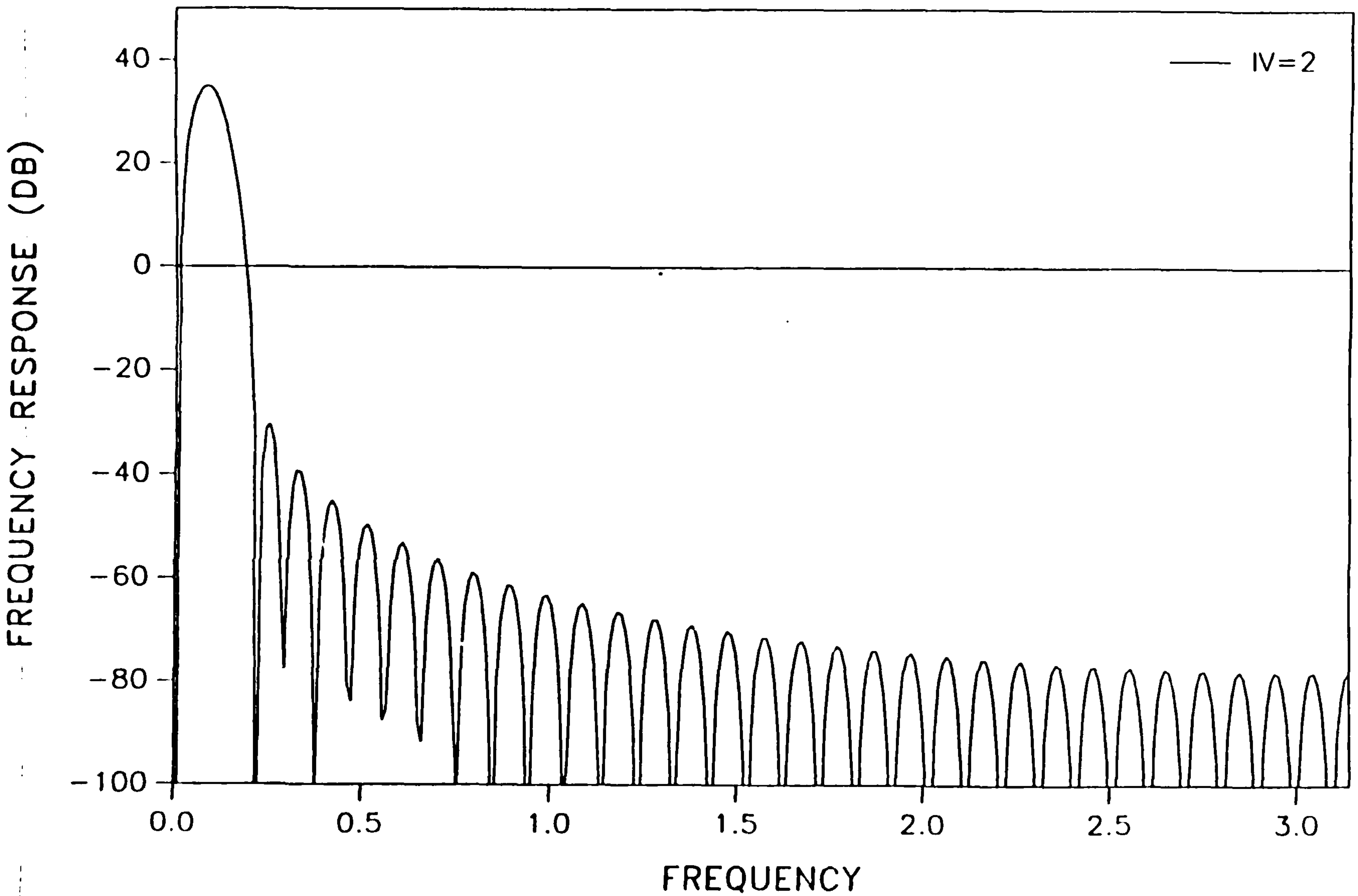
FREQUENCY SELECTIVITY OF EV–PROJECTOR

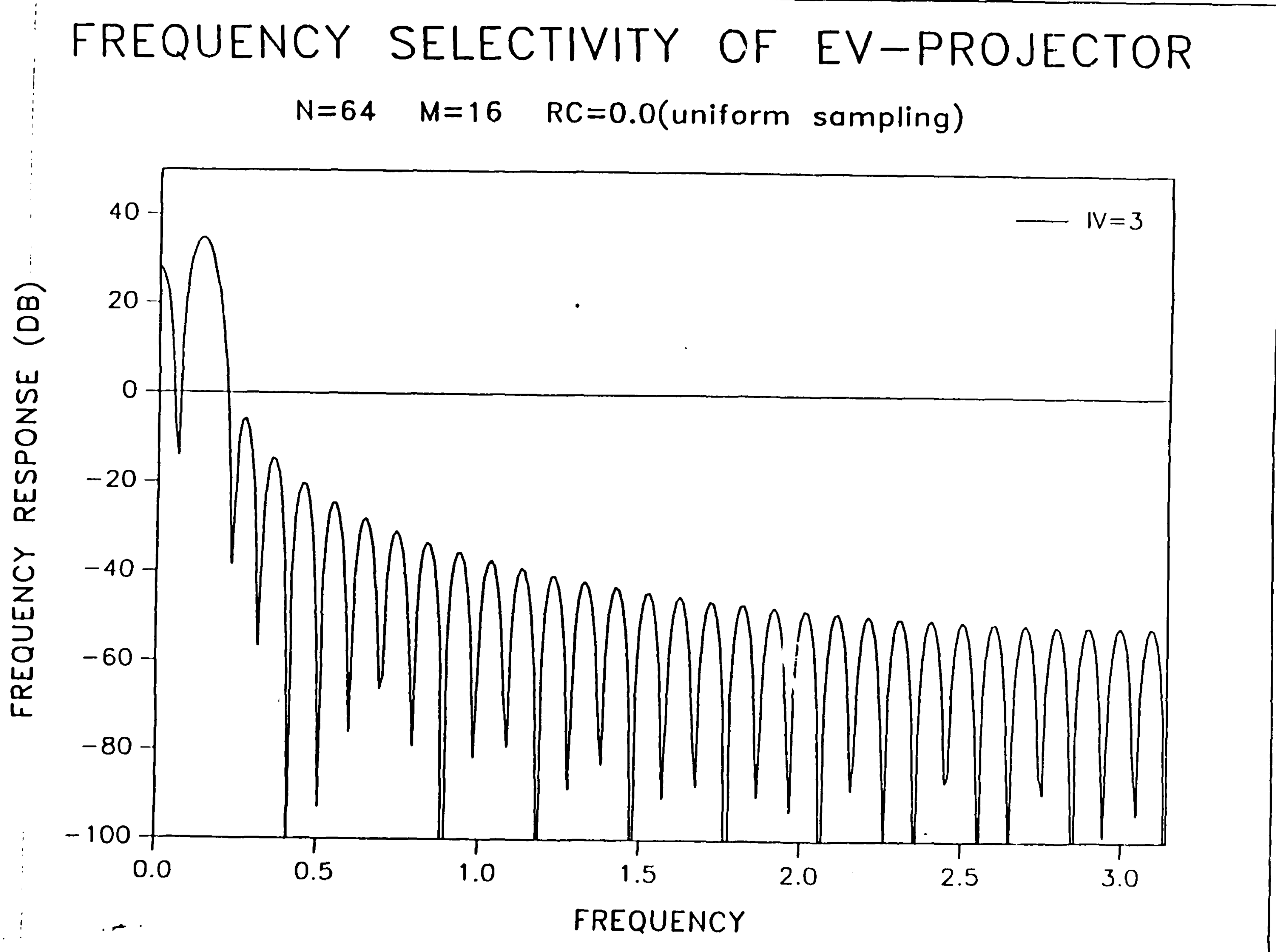
N=64 M=16 RC=0.0(uniform sampling)



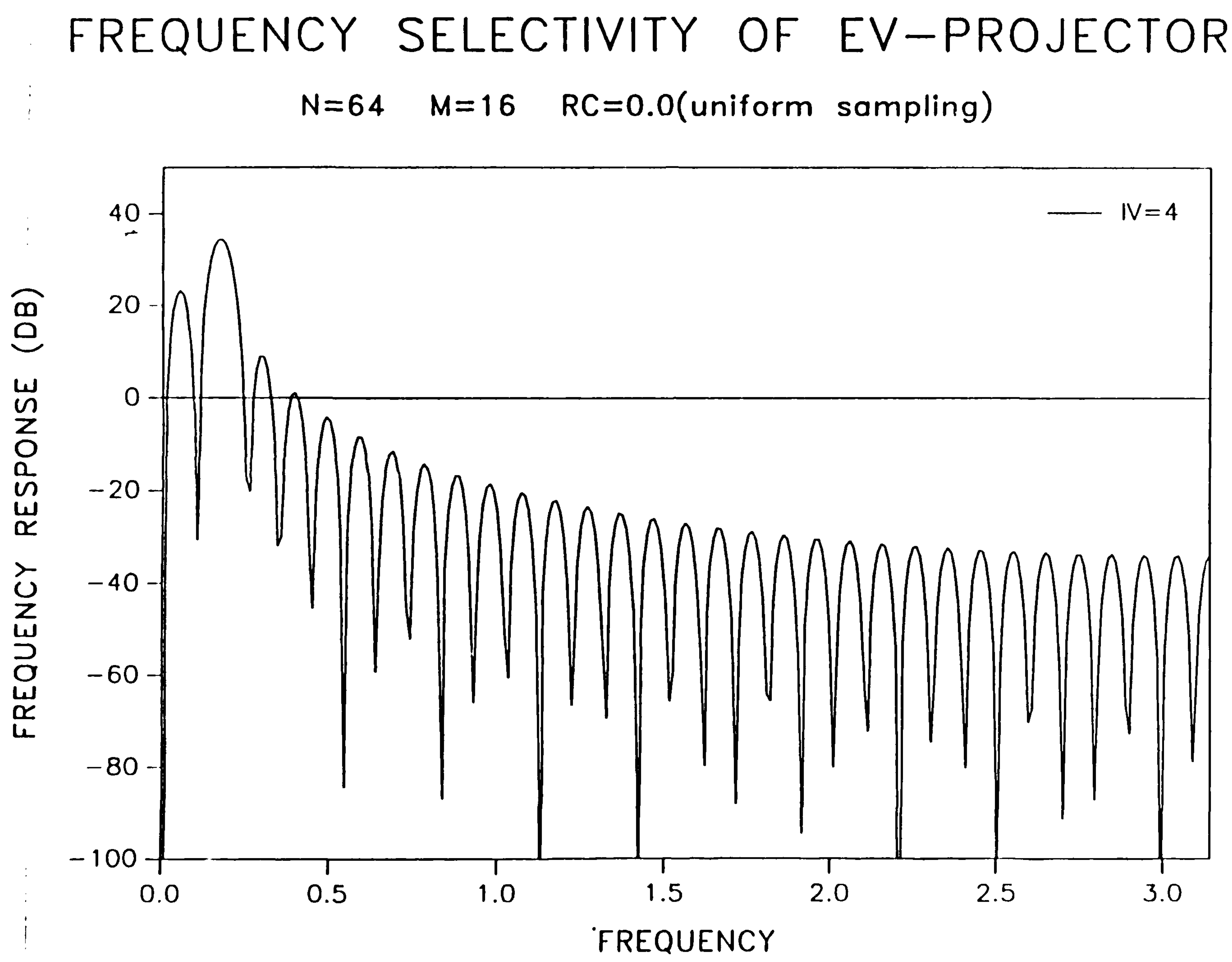
FREQUENCY SELECTIVITY OF EV–PROJECTOR

N=64 M=16 RC=0.0(uniform sampling)





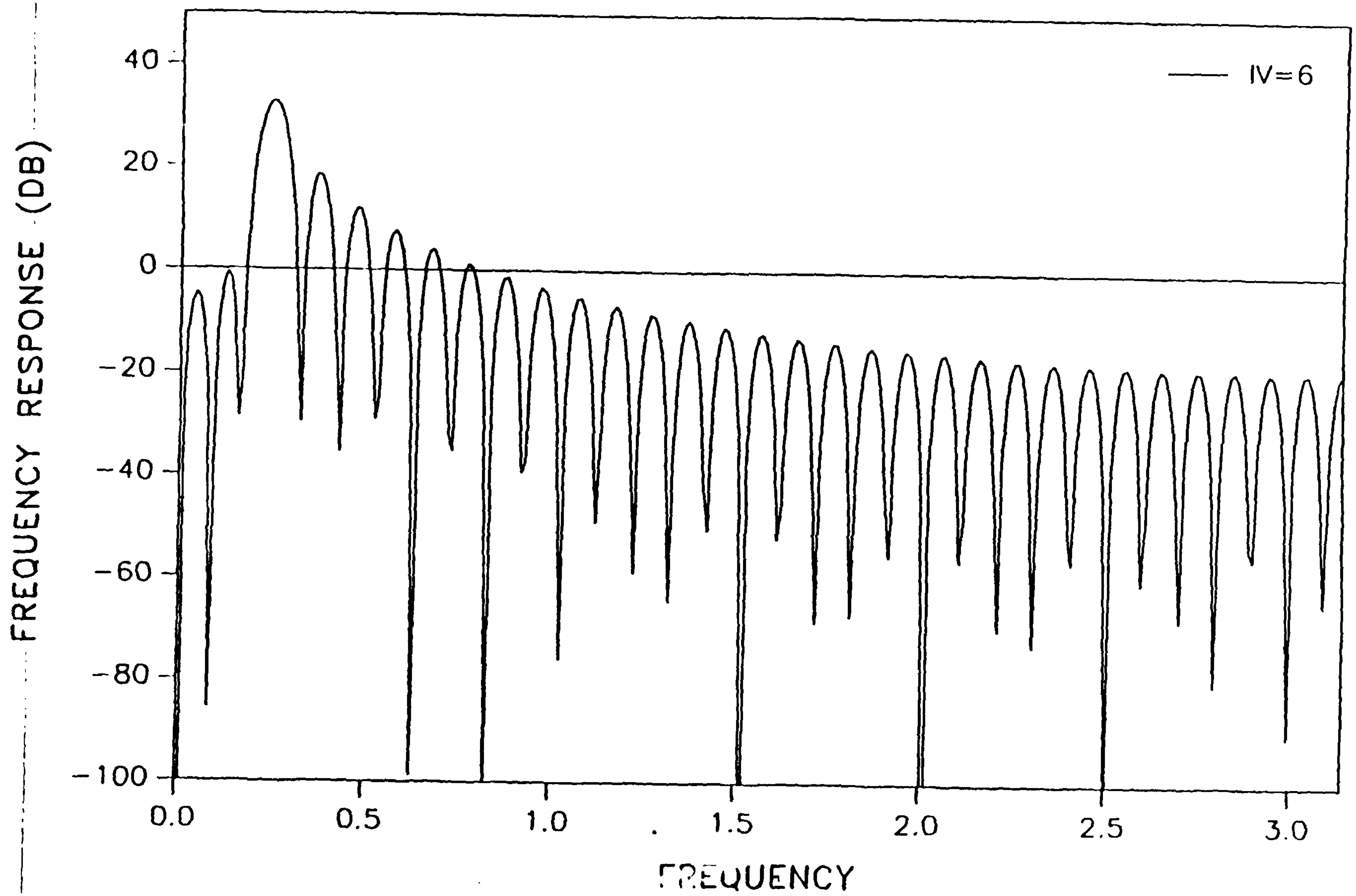
FREQUENCY SELECTIVITY OF EV-PROJECTOR
N=64 M=16 RC=0.0(uniform sampling)
IV=3
FREQUENCY RESPONSE (DB)
40
20
0
-20
-40
-60
-80
-100
0.0
0.5
1.0
1.5
2.0
2.5
3.0
FREQUENCY



FREQUENCY SELECTIVITY OF EV—PROJECTOR
N=64 M=16 RC=0.0(uniform sampling)
IV=4
FREQUENCY RESPONSE (DB)
40
20
0
-20
-40
-60
-80
-100
0.0
0.5
1.0
1.5
2.0
2.5
3.0
FREQUENCY

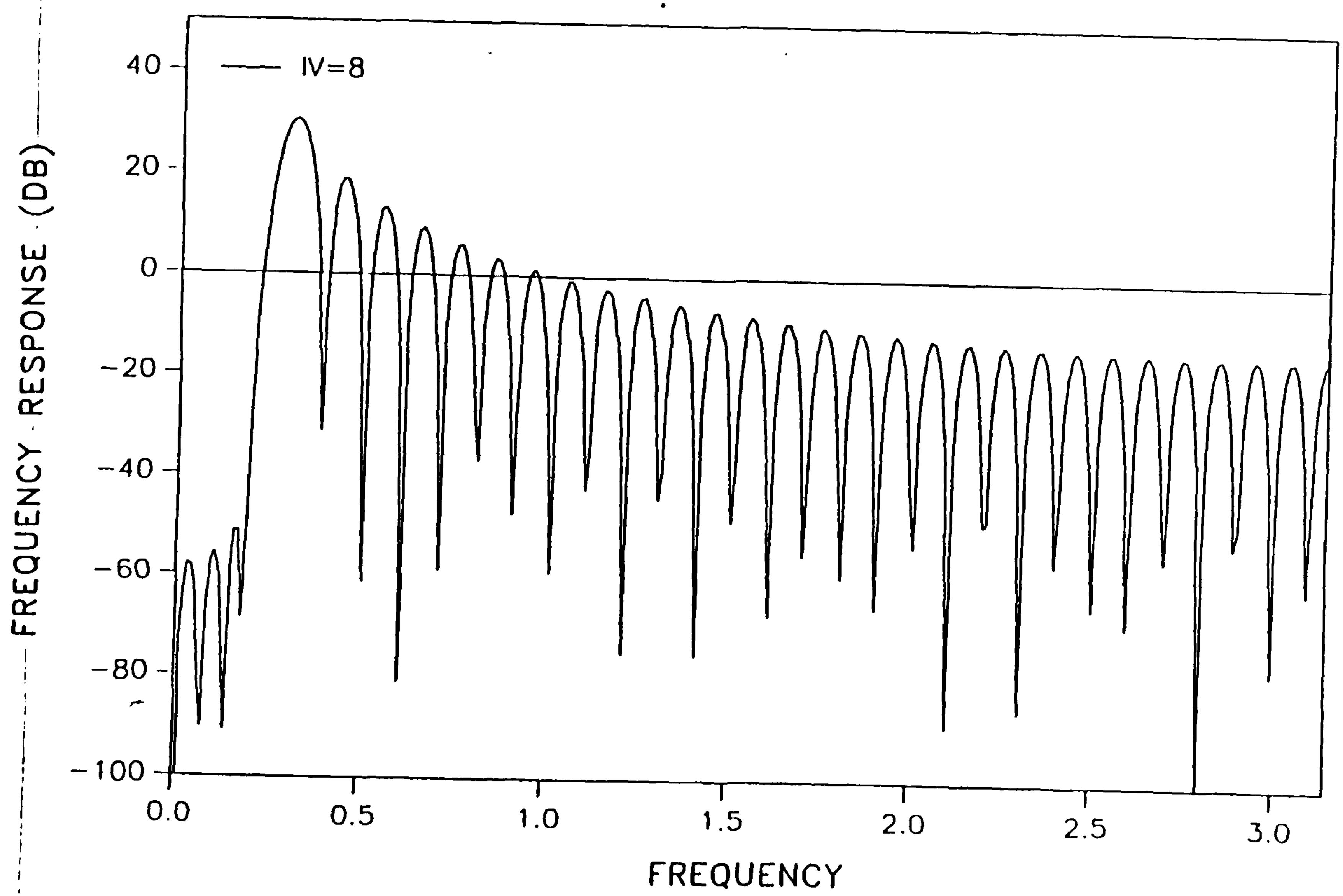
FREQUENCY SELECTIVITY OF EV-PROJECTOR

N=64   M=16   RC=0.0(uniform sampling)



FREQUENCY SELECTIVITY OF EV–PROJECTOR

N=64 M=16 RC=0.0(uniform sampling)



FREQUENCY SELECTIVITY OF EV-PROJECTOR

N=64 M=16 RC=0.0(uniform sampling)

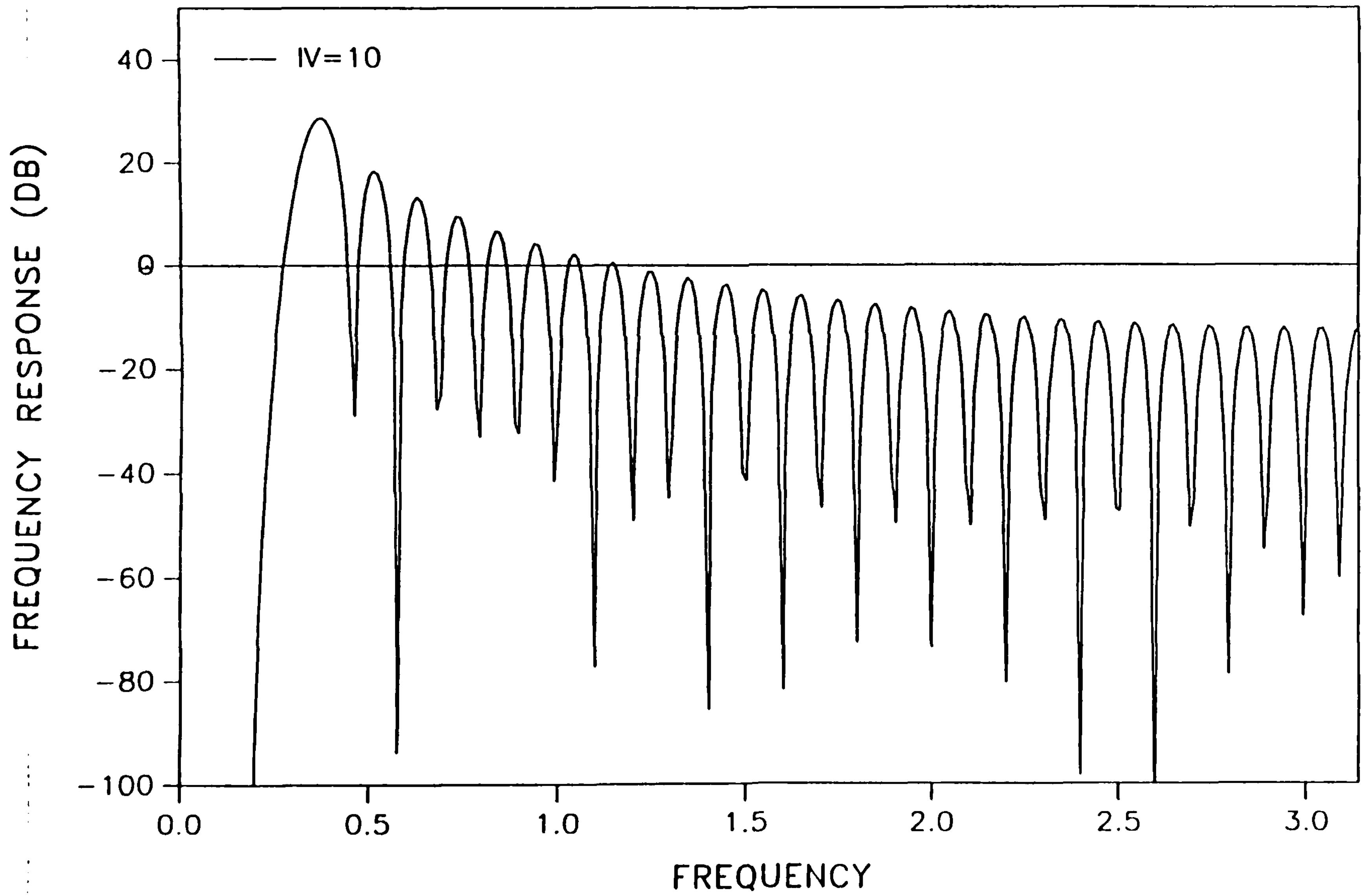
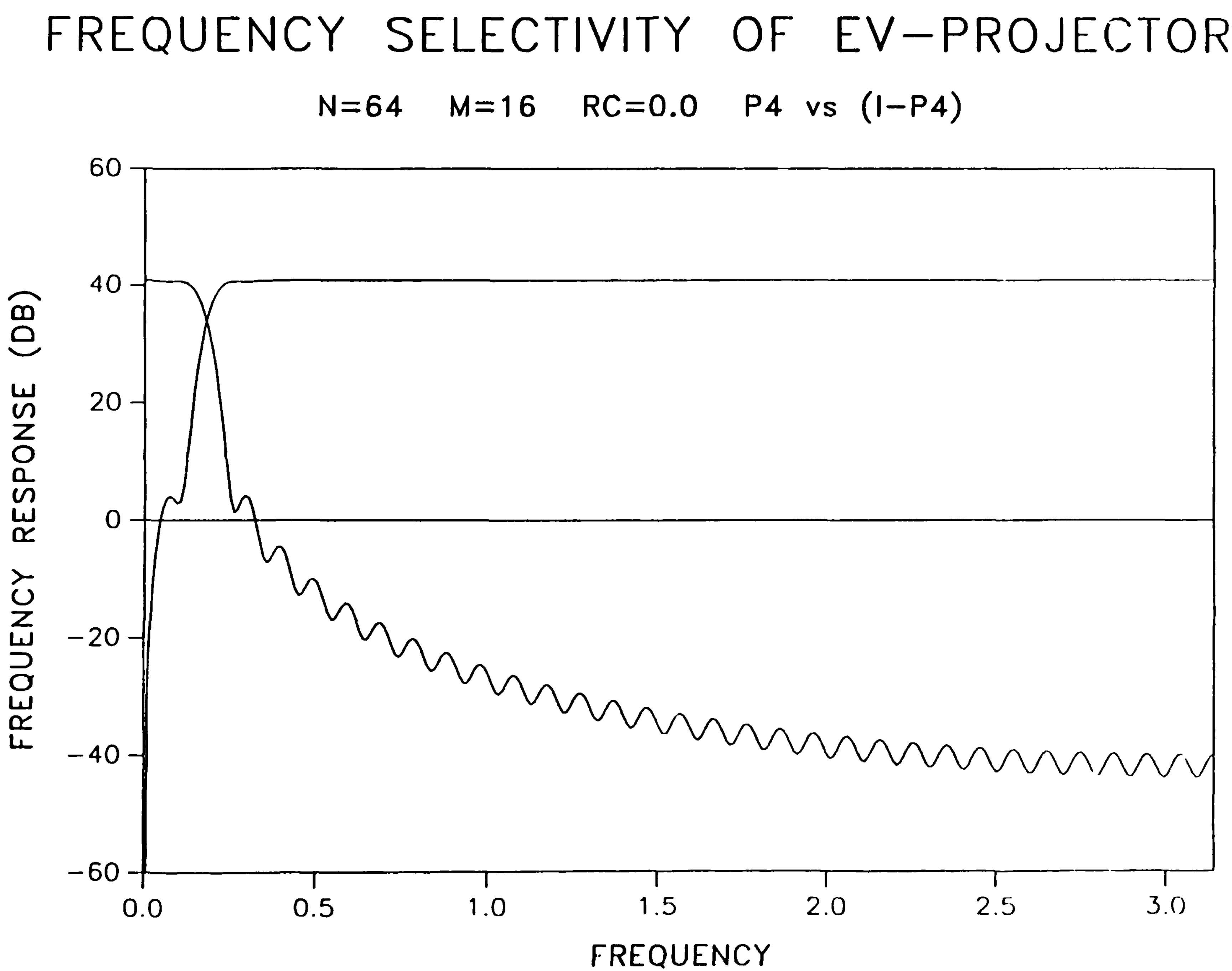


Figure 5: *Lowpass-Highpass Decomposition of Identity*


FREQUENCY SELECTIVITY OF EV-PROJECTOR
N=64 M=16 RC=0.0 P4 vs (I-P4)
FREQUENCY RESPONSE (DB)
60
40
20
0
-20
-40
-60
0.0
0.5
1.0
1.5
2.0
2.5
3.0
FREQUENCY

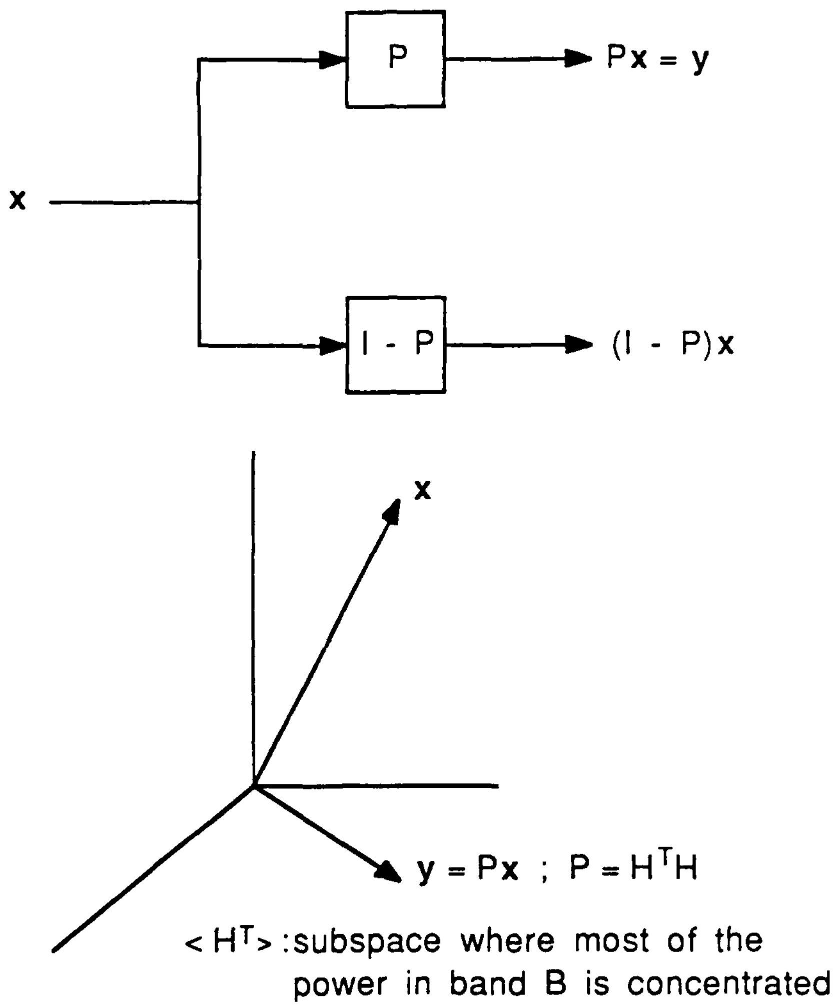


Figure 6: *The Transformation H Characterizes a Projection onto a Subspace $\langle H^T \rangle$*

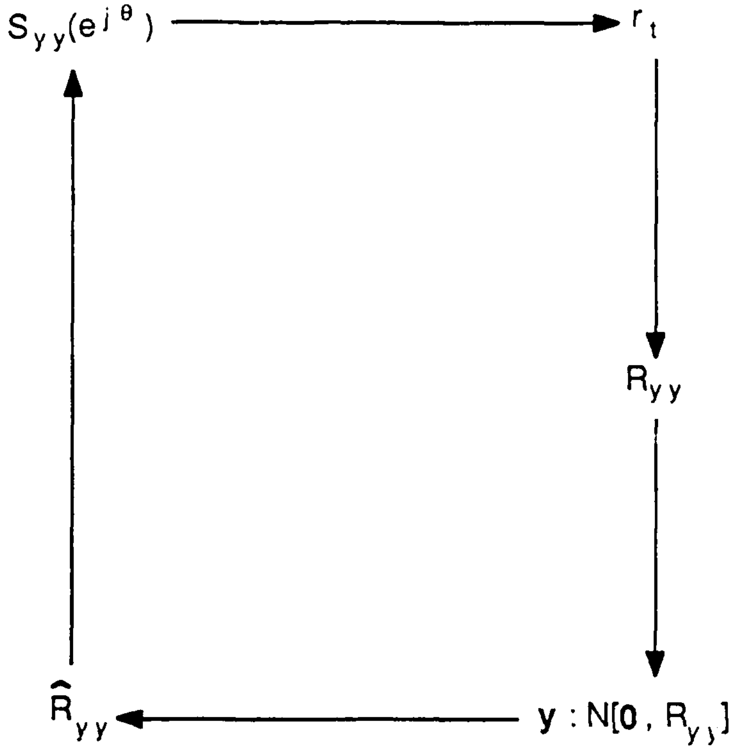
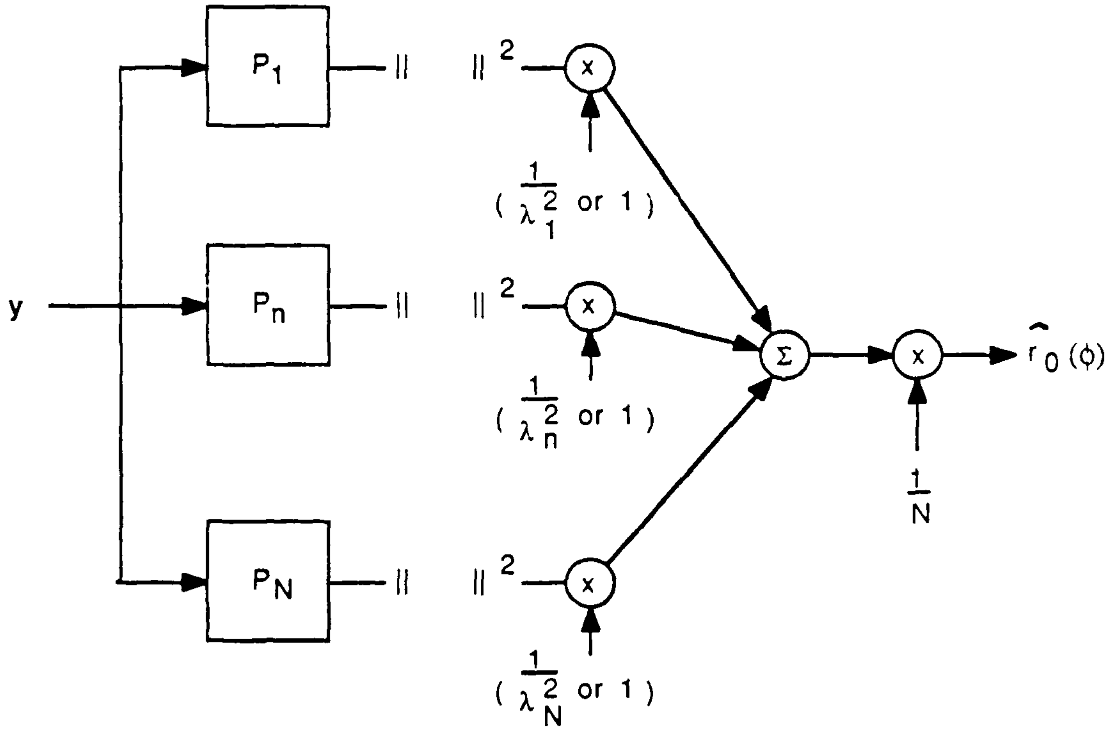


Figure 7: *Flow of ideas in Parametric Spectrum Analysis*



$$\hat{r}_0(\phi) = \text{estimate of} \int_{B(\phi)} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi}$$

Figure 8: *Maximum Likelihood Estimators of Power in Band $B(\phi)$*

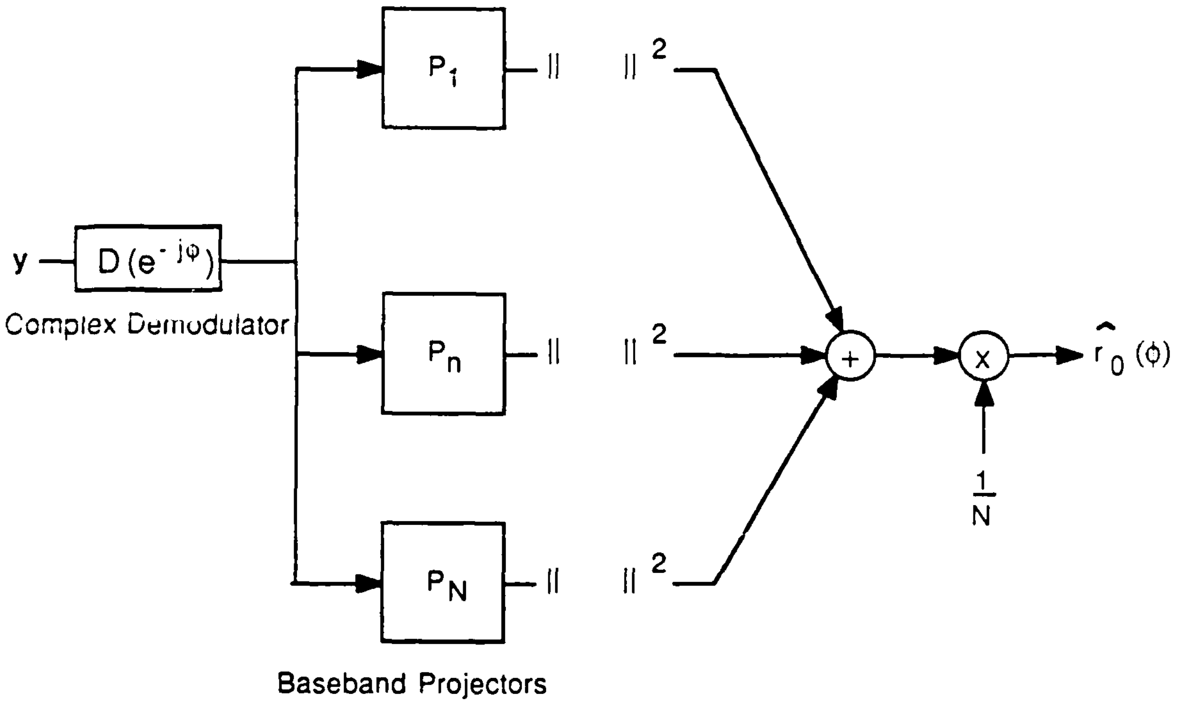


Figure 9: *Maximum Likelihood Estimation of Power Using Complex Demodulator and Baseband Projectors*

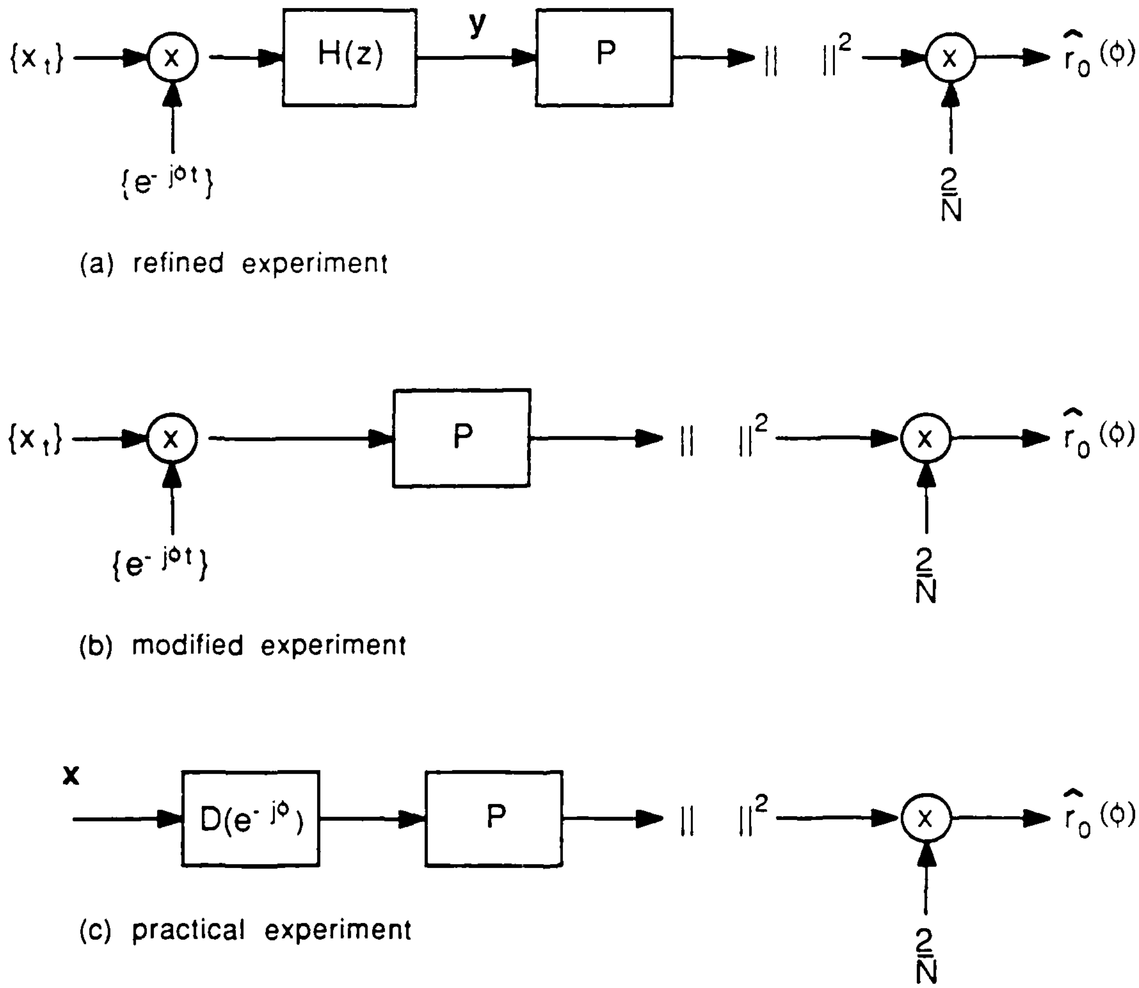


Figure 10: *Sequence of Experiments for Spectrum Analysis*

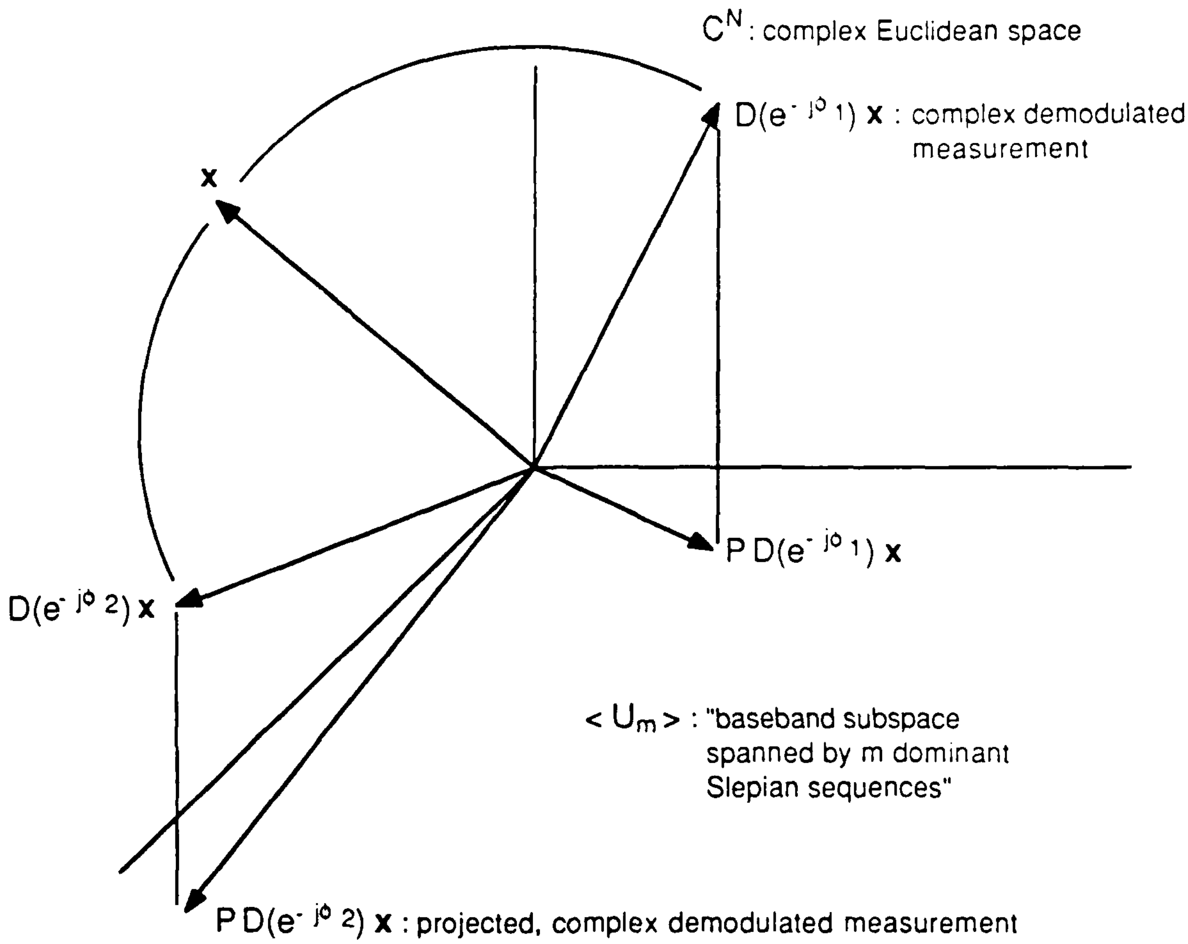
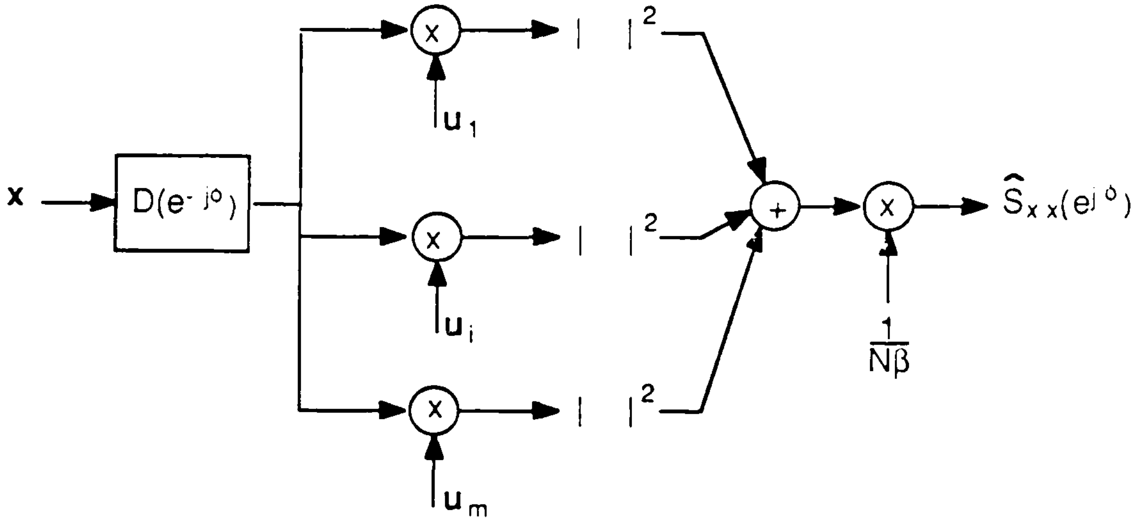
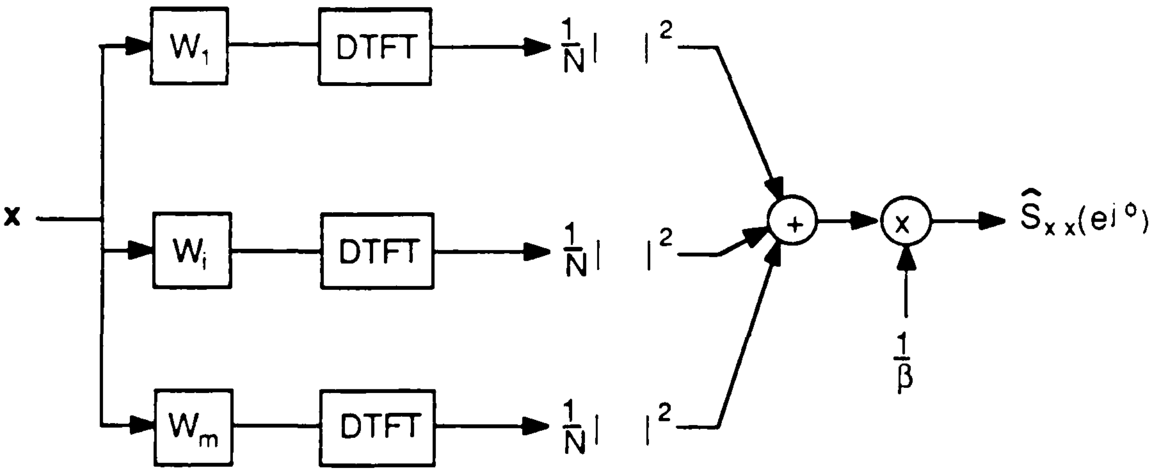


Figure 11: *Rotations and Projections for Spectrum Analysis*



(a) correlator interpretation



(b) multi-window interpretation

Figure 12: *Interpretation of Projection-Based Spectrum Estimator*

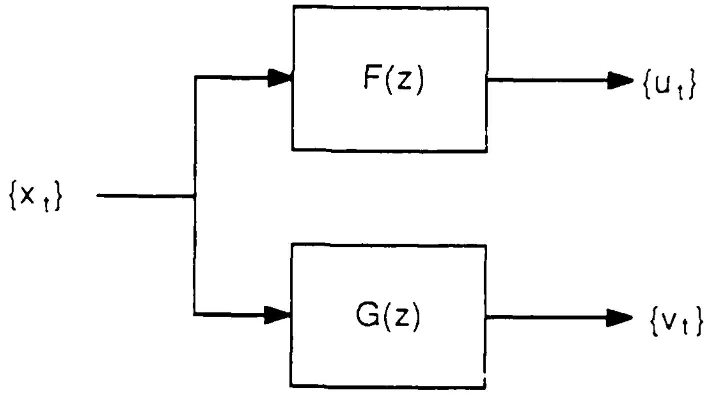


Figure 13: *More General Experiment*

Quadratic Estimators of the Power Spectrum

C. T. Mullis and Louis L. Scharf

Electrical and Computer Engineering

University of Colorado

Boulder, CO 80309

ABSTRACT

Among nonparametric estimators of the power spectrum, quadratic estimators are the only ones that are dimensionally correct. In this chapter we motivate interest in quadratic estimators by setting up idealized experiments for spectrum analysis and deriving maximum likelihood estimates for narrowband power. These idealized experiments show that quadratic functions of the experimental data are sufficients statistics for estimating power. The maximum likelihood estimates show that low-rank projection operators play a fundamental role in the theory of spectrum analysis. The projection operator plays the same role as a bandpass filter in a conventional swept frequency spectrum analyzer. The mean-squared error of a maximum likelihood estimator of power is inversely proportional to the rank of the estimator.

With our maximum likelihood results in hand, we turn to a systematic study of quadratic estimators of the power spectrum. We prove a fundamental representation theorem for any quadratic estimator of the power spectrum that is required to be positive and modulation-invariant. The resulting estimator has a multiwindow interpretation. The quadratic estimator may be tailored to produce a number of classical estimators, including those of Schuster, Daniell, Blackman and Tukey, Grenander and Rosenblatt, Clergeot, and Thomson. In one of its forms, the quadratic estimator is the maximum likelihood estimator for the power in a narrow spectral band. In this form, the estimator projects data onto a low-rank subspace where the power in the subspace is the estimate of the power in a spectral band. We complete our study of quadratic estimators by bounding their mean-squared error. The results corroborate the bounds obtained from the idealized experiments and bring a wealth of insight into the trade-off between model bias (or spectral resolution) and estimator variance.

We tie up our results by illustrating a number of equivalent implementations for quadratic forms in

low-rank projection operators.

1.0 INTRODUCTION

In the theory of wide-sense stationary time series, there is no confusion about the meaning of the term power spectrum. But what does it mean to estimate one, and what are the motivations for doing so? For Schuster [1] the motivation was to find hidden periodicities in meterological time series, and therefore it was natural for him to form a "periodogram" that would peak when periodic components of the time series matched the period of his analyzer. For Einstein [2], the motivation was to determine the variance of a Fourier series coefficient in a periodic expansion of a random waveform. Einstein's variance expression was, in fact, the power spectrum. He showed that the power spectrum was the Fourier transform of the correlation function for the random waveform. This finding, refined and extended in Wiener's work on generalized Fourier analysis [3], led to the generally held view that spectrum analysis was a problem of correlation analysis. This was the point of view adopted by Blackman and Tukey [4].

The Schuster and Einstein views are actually compatible. The periodogram and the Fourier transform of the estimated correlation sequence produce identical estimates of the power spectrum, provided the so-called biased estimate of the correlation sequence is used. The mean of the periodogram is a Bartlett-windowed version of the true spectrum, and it converges to the true spectrum as the length of the data window increases without bound. The variance of the periodogram does not converge to zero, meaning the periodogram is an inconsistent estimator of the power spectrum. The basic problem with the periodogram is that it generates roughly independent point estimates of the power spectrum at the same rate that it acquires new data, leaving no room for statistical averaging. This problem was understood by Daniell, who proposed the use of frequency averaging to stabilize the variance of an estimator of the power spectrum [5]. The closely related ideas of segmenting, windowing, and averaging, as advocated by Welch [6], are simply variations on Daniell's theme. All of these variations produce estimates of the power spectrum that are quadratic in the data.

In the sections that follow, we develop a theory for quadratic estimators of the power spectrum. We begin in Section 2.0 with a heuristic discussion of the aims and approaches of spectrum analysis. We set up an idealized experiment and illustrate the role played by bandpass filters that concentrate power in a narrow spectral band. We generalize this idea to linear transformations that concentrate power and discover Slepian sequences [7] as the appropriate basis for building bandpass subspaces. In Section 3.0 we use this idealized experiment to derive maximum likelihood estimates for the power in a narrow spectral band. The estimates are quadratic forms in the experimental data. We analyze the mean and variance of the estimators and show that the mean-squared error of a reduced rank estimator depends inversely on the rank. We then show that rank reduction is a *necessity* in any practical approach to power estimation. By introducing the idea of complex modulation, we derive Thomson's [8] family of low-rank, swept-frequency quadratic estimators of the power spectrum.

Quadratic estimators arise as sufficient statistics in an idealized experiment for measuring power in a narrow spectral band. In Section 4.0 we prove a representation theorem for quadratic estimators of the power spectrum that are non-negative and modulation-invariant. In our view, these are minimum requirements for any estimator of the power spectrum, quadratic or not. We show that the resulting class of quadratic estimators scales correctly when the data is scaled and preserves the Hermitian symmetry of the power spectrum under time reversal. We compute the mean and variance of any quadratic estimator of the power spectrum and derive a simple bound on the mean-squared error of a reduced rank quadratic estimator. The result brings insight into the tradeoff between resolution and variance and produces results much like the multi-window spectrum estimators of Thomson [8]. Our quadratic estimators subsume all of the classical windowed and smoothed estimators commonly considered for spectrum estimation, including those of Schuster [1], Daniell [5], Welch [6], Clergeot [9], and many others. In one of its more illuminating forms, the quadratic estimator matches the maximum likelihood estimator for estimating the power in a narrow spectral band. In this form the estimator projects data onto a low-rank subspace and uses the power in the subspace as an estimate of power in the narrow spectral band. The projection operator is constructed from Slepian sequences. The net result is that we have at least three interpretations of Thomson's multi-window spectrum estimator, each of which brings its own insights: (i) it is the maximum likelihood estimator of the power in a narrow spectral band; (ii) it is a projection onto a low rank subspace, followed by a power computation; and (iii) it is a reduced rank, frequency averaged periodogram (a reduced rank version of Daniell's smoothed periodogram).

We tie up our results for quadratic estimators of the power spectrum by illustrating a number of equivalent implementations. These implementations use projection operators, or their close cousins, Slepian windows, and discrete-time Fourier transforms. All of our results may be extended in a straightforward way to the estimation of power in nonuniformly sampled time series [10], [11].

2.0 HEURISTIC BEGINNINGS

Spectrum estimation may be divided into two broad categories: (i) estimation of power in a narrow spectral band, and (ii) estimation of a parametric spectrum model. The first category is often called classical spectrum analysis, and the second is often called modern spectrum analysis. One of our objectives in this chapter is to show that there are plenty of modern ideas in classical spectrum analysis.

Category (i) dominated the early work on spectrum estimation, beginning with the original work of Schuster [1] and proceeding to the current work on windowed and smoothed periodograms. In our view, this category represents the view of spectrum analysis that most closely captures the essence of the problem, namely,

> "from a finite record of a wide-sense stationary time series, estimate how the total power is distributed over narrow spectral bands."

The essential problem is to sweep a narrow-band spectrum analyzer through the Nyquist band in such a way that the band is highly resolved and the estimates of power in narrow spectral bands have low variance. Thomson's 1982 paper [8] is the definitive work on multi-window spectrum estimation. It has brought new insight into this problem and stimulated renewed interest in classical spectrum estimation.

Category (ii) has dominated the engineering literature since the publication of Burg's 1967 paper on maximum entropy and autoregressive models [12]. The key idea is to assume that the power spectrum $S_{xx}(e^{j\theta})$ belongs to some parametric class such as the autoregressive class and then to identify the parameters of the model from a record of measurements. Not surprisingly, there are connections between categories (i) and (ii). The first was established by Parzen [14] in his original advocacy of autoregressive (AR) spectrum analysis as a method to smooth rough periodograms. Parzen was also the first to recognize the importance of order selection, which is akin to rank reduction, for the control of approximation error in spectrum analysis. In Section 3.0 of this chapter, we shall establish a second connection. We shall show that the problem of estimating power in a narrow spectral band may be rephrased as a problem of estimating parameters in a structured covariance matrix. This point of view has previously been exploited indirectly by Burg et al [15] and more directly in [16]–[18]. The resulting estimator of power in a narrow spectral band is a projection-based spectrum analyzer that may be interpreted as a reduced rank, frequency averaged periodogram.

2.1 The Power Spectrum and Quadratic Estimators

The variance, or power, of a wide-sense stationary (WSS) time series $\{x_t\}$ may be written as the integral

$$r_0 = \int_{-\pi}^{\pi} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi}, \tag{1}$$

where $S_{xx}(e^{j\theta})$ is the power spectral density of the time series and $S_{xx}(e^{j\phi}) \frac{d\theta}{2\pi}$ is the infinitesimal power in the band $\{\phi - \frac{d\theta}{2} < \theta \leq \phi + \frac{d\theta}{2}\}$. This infinitesimal power is just one variance component in an infinite set of infinitely resolved variance components. It is surely unreasonable to estimate such an infinite set of variance components unless, of course, the set is smoothed or finitely parameterized. This suggests two approaches, corresponding to the classical and modern theories of spectrum analysis.

The first approach is to replace the infinitely resolved variance components $S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi}$ by the finitely resolved (or smoothed) variance components

$$r_0(\phi) = \int_{B(\phi)} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi}. \tag{2}$$

The band $B(\phi)$ is typically a passband, and $r_0(\phi)$ is the power in the band. This formulation shows spectrum analysis to be a problem in the analysis of variance (ANOVA), a problem where the estimation of infinitesimal variance components is replaced by the estimation of finite variance components.

The second approach is to use an understanding of the underlying physics to parametrically describe the spectrum, and use a statistical theory to identify the parameters. This formulation leads to the estimation of autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) power spectra. The resulting theory is closer to the theory of rational model identification (from output data) than it is to the theory of spectrum analysis.

If we choose our parametric model of the spectrum to model power in a narrow spectral band, we can bridge the gap between these two approaches.

2.2 Ideal Filters

When we represent the power of a time series by its variance components, we interpret $r_0(\phi)$ to be "that part of the total power that lies in band $B(\phi)$." But this is evidently "the power in that component of the time series that lies in band $B(\phi)$." How can we estimate this power? If we had an infinite record of data and infinite computing resources, we could pass the data (or signal) through the ideal bandpass filter illustrated in Figure 1. The frequency response of the ideal filter $H_B(z)$ is

$$H_B(e^{j\theta}) = \begin{cases} 1, & \theta \in B(\phi) \\ 0, & \theta \notin B(\phi), \end{cases} \tag{3}$$

where

$$B(\phi) = \{\theta : |\theta - \phi| \leq \beta\pi\}. \tag{4}$$

The filtered signal

$$\{y_t\} = H_B(z)\{x_t\} \tag{5}$$

is surely what we mean by the phrase "that part of the signal $\{x_t\}$ that lies in band $B(\phi)$." The variance of y_t is the power in the band:

$$\begin{aligned} r_0(\phi) = E|y_t|^2 &= \int_{B(\phi)} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi} \\ &= \sum_{t=-\infty}^{\infty} r_t \int_{B(\phi)} e^{-jt\theta} \frac{d\theta}{2\pi}. \end{aligned} \tag{6}$$

This result shows that the power in the band depends on the entire covariance sequence $\{r_t\}_{-\infty}^{\infty}$ for the time series $\{x_t\}$. If one adopts the point of view that spectrum analysis is a problem in correlation analysis, then one is forced to estimate an infinite number of correlations or to impose a parametric model for extending them outside some range of indexes for which they can be estimated.

Under an appropriate ergodic assumption, the estimator

$$\hat{r}_0(\phi) = \frac{1}{2M+1} \sum_{t=-M}^{M} |y_t|^2 \tag{7}$$

is a consistent estimator of the narrow-band power $r_0(\phi)$. This estimator may be written as the quadratic form

$$\hat{r}_0(\phi) = \sum_{n,m=-\infty}^{\infty} x_n q_{nm} x_m, \tag{8}$$

where

$$q_{nm} = \frac{1}{2M+1} \sum_{t=-M}^{M} h_{t-n} h_{t-m}. \tag{9}$$

In the limit as $M \to \infty$, the doubly indexed sequence q_{nm} depends only on $n - m$, and the quadratic form is Toeplitz. Grenander and Rosenblatt call such a Toeplitz quadratic form a spectrogram [19].

2.3 Complex Demodulation and Frequency Sweeping

When the time series $\{x_t\}$ is real, as we shall assume, then the power spectral density $S_{xx}(e^{j\theta})$ is symmetric about $\theta = 0$. The power in the narrow spectral band $B(\phi)$ is twice the power in the upper sideband $B_+(\phi)$ illustrated in Figure 2(a). But the power in this sideband is just the power in the complex demodulated signal $\{e^{-j\phi t}x_t\}$ that lies in the *baseband* $B = \{\theta : -\beta\pi < \theta \leq \beta\pi\}$. This property follows from the fact that $S_{xx}(e^{j(\theta+\phi)})$ is the power spectral density of the demodulated signal, and the power in the upper sideband may be written as

$$\begin{aligned} \frac{1}{2} r_0(\phi) &= \int_{\phi-\beta\pi}^{\phi+\beta\pi} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi} \\ &= \int_{-\beta\pi}^{\beta\pi} S_{xx}(e^{j(\theta+\phi)}) \frac{d\theta}{2\pi}. \end{aligned} \tag{10}$$

See Figure 2(b) for an illustration. This finding permits us to replace the idealized experiment of Figure 1 by the idealized experiment of Figure 3. In this experiment, the time series $\{x_t\}$ is complex demodulated in order to sweep the upper sideband down to baseband where it is filtered by the ideal baseband filter $H(z)$. The estimator

$$\frac{1}{2}\hat{r}_0(\phi) = \frac{1}{2M+1} \sum_{t=-M}^{M} |y_t|^2 \tag{11}$$

is a consistent estimator of the power in upper sideband $B_+(\phi)$ and of one-half the power in passband $B(\phi)$.

2.4 FIR Filters

The basic idea embodied in the alternative experiment of Figure 3 may be turned into a practical alternative if we replace the ideal baseband filter $H_B(z)$ with a causal, finite-dimensional FIR filter $H(z)$:

$$\begin{aligned} H(z) &= \sum_{n=0}^{N-1} h_n z^{-n} \\ &= \mathbf{h}^T \Psi(z). \end{aligned} \tag{12}$$

where

$$\Psi(z) = [1 \; z^{-1} \; z^{-2} \; \cdots \; z^{-(N-1)}]^T; \qquad \mathbf{h} = [h_0 \; h_1 \cdots h_{N-1}]^T. \tag{13}$$

If the filter $H(z)$ is to have properties analogous to those of the ideal bandpass filter, then it should be as frequency selective as we can make it. One way to measure frequency selectivity is to pass white noise through the filter and measure the output variance:

$$r_0 = \int_{-\pi}^{\pi} |H(e^{j\theta})|^2 \frac{d\theta}{2\pi} = \mathbf{h}^T\mathbf{h}. \tag{14}$$

The part of this power that lies in baseband B is

$$r_0(B) = \int_B |H(e^{j\theta})|^2 \frac{d\theta}{2\pi}. \tag{15}$$

If we use the representation $H(z) = \mathbf{h}^T\Psi(z)$, we may write this power as

$$\begin{aligned} r_0(B) &= \mathbf{h}^T \int_B \Psi(e^{j\theta})\Psi^*(e^{j\theta}) \frac{d\theta}{2\pi} \mathbf{h} \\ &= \mathbf{h}^T R \mathbf{h} \end{aligned} \tag{16}$$

where the Toeplitz matrix R depends only on the baseband B:

$$R = \int_B \Psi(e^{j\theta})\Psi^*(e^{j\theta}) \frac{d\theta}{2\pi} = \{r_{m-n}\} \tag{17}$$

$$r_{m-n} = \int_{-\beta\pi}^{\beta\pi} e^{j\theta(m-n)} \frac{d\theta}{2\pi} = \beta \operatorname{sinc}[\beta\pi(m-n)].$$

We would like to make $H(z)$ as frequency selective as we can by maximizing $r_0(B)$, under the constraint that $r_0 = \mathbf{h}^T\mathbf{h} = 1$. This leads to the following problem:

$$\max_{\mathbf{h}} \mathbf{h}^T R\mathbf{h} \quad \text{subject to the constraint } \mathbf{h}^T\mathbf{h} = 1. \tag{18}$$

The solution is to make $\mathbf{h}$ the dominant eigenvector $\mathbf{u}_1$, corresponding to the maximum eigenvalue λ_1^2, of R:

$$R = U\Lambda^2 U^T \qquad (RU = U\Lambda) \tag{19}$$

$$U = [\mathbf{u}_1\mathbf{u}_2\cdots\mathbf{u}_N] \qquad \Lambda^2 = \operatorname{diag}(\lambda_1^2\lambda_2^2\cdots\lambda_N^2) \qquad \lambda_1^2 > \lambda_2^2 > \cdots > \lambda_N^2.$$

The eigenvectors of R are the Slepian sequences featured so prominently in Thomson's paper [8].

In summary, the most frequency selective FIR filter $\mathbf{h}$ for band B is the filter

$$H(z) = \mathbf{h}^T\Psi(z), \tag{20}$$

where $\mathbf{h} = \mathbf{u}_1$ is the dominant Slepian sequence for baseband B. When $\{y_t\}$ is the output of this FIR filter, the consistent estimator

$$\hat{r}_0(\phi) = \frac{1}{2M+1} \sum_{t=-M}^{M} |y_t|^2 \tag{21}$$

is a Toeplitz quadratic form as long as $M > N$.

2.5 Linear Transformations

These results may be generalized by considering a problem that is closer yet in spirit to practical estimation of power in a spectral band. The idea is to replace the FIR filter $H(z)$ by a linear transformation H that maps an N-dimensional record of the modulated time series $\{e^{-j\phi t}x_t\}$ into an m-dimensional record of transformed measurements:

$$\mathbf{y} = HD(e^{-j\phi})\mathbf{x}, \tag{22}$$

where

$$\mathbf{y} = [y_0\ y_1 \cdots y_m]^T; \qquad \mathbf{x} = [x_0\ x_1 \cdots x_{N-1}]^T; \qquad D(e^{j\phi}) = \mathrm{diag}[1\ e^{j\phi} \cdots e^{j(N-1)\phi}] \tag{23}$$

$$H = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_m^T \end{bmatrix}; \quad m \leq N. \tag{24}$$

The frequency selectivity of the linear transformation H is measured by passing white noise $\mathbf{x}$ through it and measuring the average output variance over the m output components:

$$\begin{aligned} r_0 &= \frac{1}{m} E\|\mathbf{y}\|^2 = \frac{1}{m} E\,\mathrm{tr}[H\mathbf{x}\mathbf{x}^T H^T] \\ &= \frac{1}{m}\mathrm{tr}\Big[H \int_{-\pi}^{\pi} \Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi}\,H^T\Big]. \end{aligned} \tag{25}$$

The part of this variance (or power) that "resides in band B" is

$$\begin{aligned} r_0(B) &= \frac{1}{m}\mathrm{tr}\Big[H \int_B \Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi}\,H^T\Big] \\ &= \frac{1}{m}\mathrm{tr}[HRH^T], \end{aligned} \tag{26}$$

where R is the matrix defined in Section 2.4.

If the linear transformation is to have properties like those of the linear filter $H(z)$, then we would like it to maximize $r_0(B)$, under the constraint that each row of H have unit norm:

$$\max\ \mathrm{tr}[HRH^T] \text{ subject to the constraints } \mathbf{c}_n^T\mathbf{c}_n = 1 \qquad (n = 1, 2, \ldots, m). \tag{27}$$

The problem is to find the saddle points of the Lagrangian

$$\mathcal{L} = \sum_{n=1}^{m} \mathbf{c}_n^T R\mathbf{c}_n - \sum_{n=1}^{m} \mu_n(\mathbf{c}_n^T\mathbf{c}_n - 1). \tag{28}$$

Necessary conditions are

$$Rc_n = \lambda_n^2 c_n \tag{29}$$

for each n. This makes c_n the n^{th} dominant eigenvector of R and μ_n the corresponding eigenvalue λ_n^2. The resulting power in B is

$$r_0(B) = \frac{1}{m}\operatorname{tr}[HRH^T] = \frac{1}{m}\sum_{n=1}^{m}\lambda_n^2. \tag{30}$$

The linear transformation H may be written as a matrix of Slepian sequences:

$$H = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_m^T \end{bmatrix}. \tag{31}$$

2.6 An Illuminating Example

Let's try to illustrate the frequency selectivity of the linear transformation H. The variance r_0 may be rewritten as

$$\begin{aligned} r_0 &= \frac{1}{m}\operatorname{tr}\left[H\int_{-\pi}^{\pi}\Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi}\,H^T\right] \\ &= \frac{1}{m}\sum_{n=1}^{m}\int_{-\pi}^{\pi}\left|\mathbf{u}_n^T\Psi(e^{j\theta})\right|^2\frac{d\theta}{2\pi} \\ &= \frac{1}{m}\sum_{n=1}^{m}\int_{-\pi}^{\pi}\left|U_n(e^{j\theta})\right|^2\frac{d\theta}{2\pi} \end{aligned} \tag{32}$$

$$\left|U_n(e^{j\theta})\right|^2 = \mathbf{u}_n^T\Psi(e^{j\theta}) : \text{frequency response of the } n^{\text{th}} \text{ Slepian sequence.}$$

The component of this variance that lies in band B is

$$r_0(B) = \frac{1}{m}\sum_{n=1}^{m}\int_{B}\left|U_n(e^{j\theta})\right|^2\frac{d\theta}{2\pi}. \tag{33}$$

So, the frequency selectivity is measured by the average of the frequency selectivities for the m eigenvectors $\mathbf{u}_n$.

Let's choose $N = 64$ and $B = \{\theta : -\beta\pi < \theta < \beta\pi,\ \beta = 1/16\}$. The resulting time bandwidth product is $N\beta = 4$. The number of dominant eigenvalues of R is 4 (the time-bandwidth product). The frequency response of several dominant and subdominant Slepian sequences is illustrated in Figure 4. The frequency response of the first four eigenvectors, namely

$$\sum_{n=1}^{4} \left|U_n(e^{j\theta})\right|^2, \tag{34}$$

is plotted in Figure 5, together with the frequency response of the remaining sixty eigenvectors, namely

$$\sum_{n=5}^{64} \left|U_n(e^{j\theta})\right|^2. \tag{35}$$

These results show that the dominant eigenvectors concentrate power on the band B. The subdominant ones do little in the way of concentration. In fact, their frequency selectivity is concentrated out-of-band. The results of Figure 5 may also be interpreted as follows. The frequency response of an identity transformation is

$$\begin{aligned} N = \Psi^*(e^{j\theta})\Psi(e^{j\theta}) &= \Psi^*(e^{j\theta}) \sum_{n=1}^{N} \mathbf{u}_n \mathbf{u}_n^T \Psi(e^{j\theta}) \\ &= \Psi^*(e^{j\theta}) P \Psi(e^{j\theta}) + \Psi^*(e^{j\theta})(I - P)\Psi(e^{j\theta}), \end{aligned} \tag{36}$$

where P is the rank-m projection built from the dominant Slepian sequences. The frequency response of the projection is just the frequency response of the dominant eigenvectors:

$$\Psi^*(e^{j\theta}) P \Psi(e^{j\theta}) = \sum_{n=1}^{m} \left|U_n(e^{j\theta})\right|^2, \tag{37}$$

where

$$P = \sum_{n=1}^{m} \mathbf{u}_n \mathbf{u}_n^T = H^T H. \tag{38}$$

It is clear that the linear transformation H characterizes the projection P, whose range is a low-rank subspace where most of the power in the band B is concentrated. This is illustrated in Figure 6.

2.7 Maximum Entropy Interpretation

There is another way to describe the $m \times N$ linear transformation H derived in Section 2.5. It is a linear transformation that produces a maximum entropy version of white noise, under the constraint that the rows of H are normalized. The resulting entropy of $\mathbf{y} = H\mathbf{x}$ when $\mathbf{x}$ is white is

$$\begin{aligned} \mathcal{E} &= -\ln\left[(2\pi e)^{-m/2}(\det HH^T)^{-1/2}\right] \\ &= \frac{m}{2}\ln(2\pi e). \end{aligned} \tag{39}$$

The random vector $\mathbf{y} = H\mathbf{x}$ is also as white as it can be, with covariance

$$R_{yy} = E\,\mathbf{y}\mathbf{y}^T = E\,H\mathbf{x}\mathbf{x}^T H^T = I, \tag{40}$$

and its prediction error is as large as it can be.

The norm-squared of $\mathbf{y}$ may be written as a quadratic form in the projection operator P:

$$\begin{aligned} \mathbf{y}^T\mathbf{y} &= \mathbf{x}^T H^T H\mathbf{x} \\ &= \mathbf{x}^T P\mathbf{x} = \|P\mathbf{x}\|^2. \end{aligned} \tag{41}$$

The linear transformation H that maximizes entropy also characterizes the m-dimensional subspace $\langle H^T \rangle$ where the largest fraction of the power in band B lies (and vice versa). This power is approximately the integral of the m-dominant frequency responses $\left|U_n(e^{j\theta})\right|^2$ over the band B:

$$\begin{aligned} E\,\mathbf{y}^T\mathbf{y} &= \mathrm{tr}[P] \\ &= \sum_{n=1}^{m} \|\mathbf{u}_n\|^2 = \sum_{n=1}^{m} \int_{-\pi}^{\pi} \left|U_n(e^{j\theta})\right|^2 \frac{d\theta}{2\pi} \\ &\cong \sum_{n=1}^{m} \int_{B} \left|U_n(e^{j\theta})\right|^2 \frac{d\theta}{2\pi}. \end{aligned} \tag{42}$$

3.0 MAXIMUM LIKELIHOOD ESTIMATION OF POWER

We shall begin our study of maximum likelihood theory, and its application to spectrum analysis, by returning to the experimental setup illustrated in Figure 1. However, we now ask what happens when only the finite snapshot $\mathbf{y} = [y_0\ y_1 \cdots y_{N-1}]^T$ is available to the experimenter.

Let us assume that the time series $\{x_t\}$ is Gaussian and wide-sense stationary (WSS), with mean zero, covariance sequence $\{r_t\}$, and power spectrum $S_{xx}(e^{j\theta})$. The snapshot $\mathbf{y}$ is then a normal random vector. Its mean is zero and its covariance matrix R_{yy} is related to the power spectrum of $\{x_t\}$ as follows:

$$R_{yy} = E\,\mathbf{y}\mathbf{y}^T = \int_{B(\phi)} S_{xx}(e^{j\theta})\Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi}. \tag{43}$$

The vector $\Psi(e^{j\theta})$ is called a Fourier vector; when evaluated at $\theta = 2\pi m/N$, $m = 0, 1, \ldots, N-1$, it is called a DFT vector.

Our shorthand notation for describing the snapshot $\mathbf{y}$ is

$$\mathbf{y} : N[0, R_{yy}]. \tag{44}$$

It is clear that the information about the power in band $B(\phi)$ can be carried only in the covariance structure of $\mathbf{y}$. The covariance matrix R_{yy} is Toeplitz, meaning that the power in band $B(\phi)$ is related to R_{yy} as follows:

$$r_0(\phi) = \frac{1}{N}\operatorname{tr}[R_{yy}] = \int_{B(\phi)} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi}. \tag{45}$$

The problem of estimating the power $r_0(\phi)$ in the band $B(\phi)$ is a problem in estimating the trace of R_{yy}. In maximum likelihood theory, the invariance theorem shows that the problem of estimating $\operatorname{tr} R_{yy}$ is a problem of estimating R_{yy} and then finding its trace. This brings us to a discussion of maximum likelihood estimation of covariance matrices. In the discussions to follow, we replace $r_0(\phi)$ with the notation r_0 in order to simplify notation. But r_0 is always $r_0(\phi)$, and estimates $\hat{r}_0$ are always estimates $\hat{r}_0(\phi)$.

3.1 Maximum Likelihood Estimation of the Covariance Matrix

The flow of ideas for parametric spectrum analysis is illustrated in Figure 7. In the figure, $S_{yy}(e^{j\theta})$ and r_t denote the spectrum and correlation sequence of $\{y_t\}$. If the measurements are drawn from this wide-sense stationary time series, then the covariance matrix for the measurements may be constructed from the covariance sequence. The covariance matrix is denoted R_{yy}. If the measurement sequence $\mathbf{y}$ is multivariate normal, denoted $\mathbf{y} : \mathcal{N}[0, R_{yy}]$, then the covariance matrix R_{yy} may be estimated from a criterion like maximum likelihood. The estimated covariance matrix is then used to estimate the spectrum. Generally, the covariance matrix R_{yy} has some special structure, with unknown parameters. Procedures such as maximum likelihood for identifying the parameters are generally very complicated because the parameters are imbedded in the determinant and inverse of a covariance matrix. However, if the time series is autoregressive, linear prediction approximately solves the maximum likelihood problem [13].

In the theory of maximum likelihood (ML), information about the covariance matrix is carried in the log likelihood function $L(R_{yy};\mathbf{y})$:

$$L(R_{yy};\mathbf{y}) = \frac{-N}{2}\ln(2\pi) - \frac{1}{2}\ln\det R_{yy} - \frac{1}{2}\mathbf{y}^T R_{yy}^{-1}\mathbf{y}. \tag{46}$$

The structured covariance matrix R_{yy} we denote by $R_{yy}(\underline{\theta})$, where $\underline{\theta}$ is a set of parameters $\underline{\theta} = (\theta_1\theta_2\cdots\theta_p)$. The maximum likelihood equation for determining the maximum likelihood estimate $\hat{R}_{yy} = R_{yy}(\hat{\underline{\theta}})$ is [15]

$$\mathrm{tr}\left[R_{yy}^{-1}(R_{yy} - \mathbf{y}\mathbf{y}^T)\frac{\partial R_{yy}}{\partial\theta_n}\right] = 0; \qquad (n = 1, 2, \ldots, p). \tag{47}$$

The Fisher information matrix is

$$J = \{J_{mn}\} \tag{48}$$

$$J_{mn} = \frac{1}{2}\mathrm{tr}\left[R_{yy}^{-1}\frac{\partial R_{yy}}{\partial\theta_m}R_{yy}^{-1}\frac{\partial R_{yy}}{\partial\theta_n}\right],$$

and the Cramer-Rao bound on any unbiased estimator of $\underline{\theta}$ is

$$M = E(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})^T \geq J^{-1}. \tag{49}$$

Any estimator that achieves the bound is efficient and minimum variance unbiased (MVUB).

In the following section we show how to apply these general results to the problem of estimating power in a narrow spectral band.

3.2 Unstructured Covariance

When no structural information about the covariance matrix R_{yy} is known, then the ML estimate of R_{yy} is

$$\hat{R}_{yy} = \mathbf{y}\mathbf{y}^T, \tag{50}$$

and the corresponding ML estimate of the power in band $B(\phi)$ is

$$\begin{aligned}\hat{r}_0 &= \frac{1}{N}\,\mathrm{tr}[\mathbf{y}\mathbf{y}^T] \\ &= \frac{1}{N}\,\mathbf{y}^T\mathbf{y}.\end{aligned} \tag{51}$$

The mean and variance of the estimator are

$$\begin{aligned}E\hat{r}_0 &= \frac{1}{N}\,\mathrm{tr}[E\,\mathbf{y}\mathbf{y}^T] = \frac{1}{N}\,\mathrm{tr}[R_{yy}] \\ &= \int_{B(\phi)} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi}\end{aligned} \tag{52}$$

$$\mathrm{var}\,\hat{r}_0 = \frac{2}{N^2}\,\mathrm{tr}[R_{yy}^2]. \tag{53}$$

This is a very primitive estimator of the power that exploits none of the prior information that we have about R_{yy}. For example, we know that R_{yy} is Toeplitz, and this information can be used [15]. But more than this, we know that R_{yy} should have a special spectral representation that it inherits from the ideal bandpass filter $H_B(e^{j\theta})$. This is the idea that we exploit in the next section.

3.3 Structured Covariance

The real, symmetric, non-negative definite covariance matrix R_{yy} has the spectral, or orthogonal, representation

$$R_{yy} = U\Sigma^2U^T; \qquad U^TU = I \tag{54}$$

$$U = [\mathbf{u}_1\mathbf{u}_2\cdots\mathbf{u}_N]; \qquad \Sigma^2 = \mathrm{diag}(\sigma_1^2\sigma_2^2\cdots\sigma_N^2). \tag{55}$$

We shall assume that U is a known orthogonal matrix and Σ^2 is an unknown diagonal matrix. We shall represent the diagonal matrix Σ^2 as

$$\Sigma^2 = r_0\Lambda^2; \qquad \Lambda^2 = \mathrm{diag}(\lambda_1^2\lambda_2^2\cdots\lambda_N^2) \tag{56}$$

$$\frac{1}{N}\sum_{n=1}^{N}\lambda_n^2 = 1. \tag{57}$$

With this representation of R_{yy}, the unknown parameters are r_0, the power in the band, and the nuisance parameters $\{\lambda_n^2\}_1^N$. The normalization of the parameters $\{\lambda_n^2\}_1^N$ preserves the connection between r_0 and $\frac{1}{N}\,\mathrm{tr}[R_{yy}]$:

$$\begin{aligned}r_0 &= \frac{1}{N}\,\mathrm{tr}[R_{yy}] = r_0\,\frac{1}{N}\,\mathrm{tr}[\Lambda^2] \\ &= r_0.\end{aligned} \tag{58}$$

The log-likelihood function takes the very special form

$$L(r_0, \{\lambda_n^2\}_1^N; \mathbf{y}) = \frac{-N}{2} \ln(2\pi) - \frac{1}{2} \ln \prod_{n=1}^{N} r_0 \lambda_n^2 - \frac{1}{2} \sum_{n=1}^{N} \frac{1}{r_0 \lambda_n^2} \mathbf{y}_n^T P_n \mathbf{y}_n. \tag{59}$$

In this likelihood formula, P_n is a rank one projection onto the subspace spanned by the n^{th} eigenvector $\mathbf{u}_n$ and $\mathbf{y}_n^T P_n \mathbf{y}_n$ is one of N sufficient statistics:

$$\mathbf{y}_n^T P_n \mathbf{y}_n : \text{sufficient} \tag{60}$$

$$P_n = \mathbf{u}_n \mathbf{u}_n^T.$$

This finding is important because it shows that, whether or not we use a ML theory, the quadratic forms $\mathbf{y}_n^T P_n \mathbf{y}_n$ are sufficient statistics for the parameters r_0, $\{\lambda_n^2\}_1^N$.

Let's maximize likelihood under the constraint that the $\{\lambda_n^2\}_1^N$ average to one. The appropriate Lagrangian is

$$\mathcal{L} = -\sum_{n=1}^{N} \ln(r_0 \lambda_n^2) - \sum_{n=1}^{N} \frac{1}{r_0 \lambda_n^2} \mathbf{y}_n^T P_n \mathbf{y} - \mu \Big(\frac{1}{N} \sum_{n=1}^{N} \lambda_n^2 - 1 \Big). \tag{61}$$

The gradients of $\mathcal{L}$ with respect to r_0 and λ_n^2 may be equated to zero to produce the ML equations

$$\frac{\partial \mathcal{L}}{\partial r_0} = 0 : \frac{1}{r_0} \sum_{n=1}^{N} \frac{1}{\lambda_n^2} \mathbf{y}^T P_n \mathbf{y} - N = 0 \tag{62}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_n^2} = 0 &: \frac{r_0}{(r_0 \lambda_n^2)^2} \mathbf{y}^T P_n \mathbf{y} - \frac{1}{\lambda_n^2} - \frac{\mu}{N} = 0 \\ &: \frac{1}{r_0} \frac{1}{\lambda_n^2} \mathbf{y}^T P_n \mathbf{y} - 1 - \frac{\mu}{N} \lambda_n^2 = 0. \end{aligned} \tag{63}$$

The constraint may be invoked in the second of these equations to show that the constraint plays no role in the minimization:

$$\begin{gathered} \frac{1}{r_0} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\lambda_n^2} \mathbf{y}^T P_n \mathbf{y} - 1 - \frac{\mu}{N} \frac{1}{N} \sum_{n=1}^{N} \lambda_n^2 = 0 \\ 1 - 1 - \frac{\mu}{N} = 0 \implies \mu = 0. \end{gathered} \tag{64}$$

The maximum likelihood estimates of r_0 and λ_n^2 are

$$\begin{gathered} \hat{r}_0 = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\hat{\lambda}_n^2} \mathbf{y}^T P_n \mathbf{y} \\ \hat{\lambda}_n^2 = \frac{1}{\hat{r}_0} \mathbf{y}^T P_n \mathbf{y}. \end{gathered} \tag{65}$$

There are actually two important results here. First, if the $\{\lambda_n^2\}_1^N$ are known, then the ML estimate of r_0 is

$$\hat{r}_0 = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\lambda_n^2}\mathbf{y}^T P_n \mathbf{y}. \tag{66}$$

Second, if the $\{\lambda_n^2\}_1^N$ are unknown, then the normalizing equation may be used to rewrite $\hat{r}_0$ as

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\lambda}_n^2 = 1 = \frac{1}{\hat{r}_0}\,\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}^T P_n \mathbf{y} \tag{67}$$

$$\hat{r}_0 = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}^T P_n \mathbf{y} = \frac{1}{N}\mathbf{y}^T\mathbf{y}.$$

These two findings, illustrated in Figure 8, show that the ML estimate of the power in band $B(\phi)$ is quadratic in the data.

3.4 Mean and Variance

The mean and variance of our two estimators are computed from the mean and variance of the quadratic, sufficient, statistics $\mathbf{y}^T P_n \mathbf{y}$. The rank-one projection $P_n\mathbf{y}$ is distributed as

$$P_n\mathbf{y} : N[\mathbf{0}, r_0\lambda_n^2\Delta_n] \tag{68}$$

$$\Delta_n = \mathrm{diag}(0\cdots 0\ 1\ 0\cdots 0).$$

This makes the quadratic form $\mathbf{y}^T P_n \mathbf{y} = ||P_n\mathbf{y}||^2$ a scaled chi-squared random variable with the following mean and variance:

$$E\,\mathbf{y}^T P_n \mathbf{y} = r_0\lambda_n^2 \tag{69}$$

$$\mathrm{var}[\mathbf{y}^T P_n \mathbf{y}] = 2(r_0\lambda_n^2)^2. \tag{70}$$

It is easy to see that the sufficient statistics are uncorrelated and therefore independent:

$$\begin{aligned} E(P_n\mathbf{y})(P_m\mathbf{y})^T &= P_n R P_m \\ &= \lambda_n^2 \mathbf{u}_n \mathbf{u}_n^T \delta_{mn}. \end{aligned} \tag{71}$$

This means that we may add variances at will to compute the variance of $\hat{r}_0$.

Unknown $\{\lambda_n^2\}_1^N$. When the eigenvalues of R are unknown, then the ML estimate of r_0 is

$$\hat{r}_0 = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}^T P_n \mathbf{y}. \tag{72}$$

The mean and variance of this estimator are

$$E\hat{r}_0 = \frac{1}{N}\sum_{n=1}^{N} r_0\lambda_n^2 = r_0 \tag{73}$$

$$\operatorname{var}\hat{r}_0 = \frac{1}{N^2}\sum_{n=1}^{N} 2(r_0\lambda_n^2)^2 = \frac{2r_0^2}{N}\,\frac{1}{N}\sum_{n=1}^{N}(\lambda_n^2)^2. \tag{74}$$

It is interesting to ask which distribution of the unknown $\{\lambda_n^2\}_1^N$ minimizes the variance, under the constraint that $\frac{1}{N}\sum_{n=1}^{N}\lambda_n^2 = 1$. The obvious answer is $\lambda_n^2 = 1$ for all n, in which case

$$\operatorname{var}\hat{r}_0 = \frac{2r_0^2}{N}. \tag{75}$$

This question is much more fascinating, and the answer more profound, when the estimator is replaced by a low-rank approximant. We take up this question in Section 3.5.

Known $\{\lambda_n^2\}$**.** When the eigenvalues of R are known up to the scale constant r_0, then the ML estimate of r_0 is

$$\hat{r}_0 = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\lambda_n^2}\mathbf{y}^T P_n \mathbf{y}. \tag{76}$$

The mean and variance of the estimator are

$$E\hat{r}_0 = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\lambda_n^2}(r_0\lambda_n^2) = r_0 \tag{77}$$

$$\operatorname{var}\hat{r}_0 = \frac{1}{N^2}\sum_{n=1}^{n}\frac{1}{(\lambda_n^2)^2}\,2(r_0\lambda_n^2)^2 = \frac{2r_0^2}{N}. \tag{78}$$

This latter finding is very important because it establishes *the* lower bound on the variance of any unbiased estimator of the power in band $B(\phi)$. Why do we say this? Because the idealized experimental setup illustrated in Figure 1 preserves all information about the power in the band, and the snapshot contains all of the information that one can get in a finite snapshot. The assumption that the eigenstructure of R is known up to a multiplicative constant in the eigenvalues represents as much apriori knowledge as one can have about the spectrum $S_{xx}(e^{j\theta})$ without giving away its power in the band. Without prior knowledge of the normalized eigenvalues $\{\lambda_n\}_1^N$, the minimizing distribution of eigenvalues produces the same minimum variance. Thus, for any *unbiased* estimator of the power in the band,

$$\operatorname{var}\Big(\int_{B(\phi)} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi}\Big) \geq \frac{2}{N}\Big(\int_{B(\phi)} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi}\Big)^2. \tag{79}$$

3.5 Reduced Rank Estimators

The estimators we have so far derived are defective in the sense that they require detailed knowledge of eigenstructure for their implementation. Typically, such knowledge is hard to come by because eigenstructure for subdominant spectral modes is notoriously unreliable. So, what if we replace our ML estimators by low-rank approximations of them?

Unknown $\{\lambda_n^2\}_1^N$. The obvious low-rank approximation of $\hat{r}_0$ is

$$\dot{r}_0 = \frac{1}{N}\sum_{n=1}^{m} \mathbf{y}^T P_n \mathbf{y}. \tag{80}$$

The mean and variance of the low-rank estimate are

$$E\,\dot{r}_0 = \frac{1}{N}\sum_{n=1}^{m}(r_0\lambda_n^2) = r_0\,\frac{1}{N}\sum_{n=1}^{m}\lambda_n^2 \tag{81}$$

$$\operatorname{var}\dot{r}_0 = \frac{1}{N^2}\sum_{n=1}^{m} 2(r_0\lambda_n^2)^2 = \frac{2r_0^2}{N}\,\frac{1}{N}\sum_{n=1}^{m}(\lambda_n^2)^2. \tag{82}$$

The mean-squared error of this reduced rank estimator is

$$\begin{aligned}
\mathrm{MSE}(m) &= E(\dot{r}_0 - r_0)^2 = (E\,\dot{r}_0 - r_0)^2 + \operatorname{var}\dot{r}_0 \\
&= \left(r_0\,\frac{1}{N}\sum_{n=1}^{m}\lambda_n^2 - r_0\,\frac{1}{N}\sum_{n=1}^{N}\lambda_n^2\right)^2 + \frac{2r_0^2}{N}\,\frac{1}{N}\sum_{n=1}^{m}(\lambda_n^2)^2 \\
&= \frac{r_0^2}{N^2}\left(\sum_{n=m+1}^{N}\lambda_n^2\right)^2 + \frac{2r_0^2}{N^2}\sum_{n=1}^{m}(\lambda_n^2)^2.
\end{aligned} \tag{83}$$

This mean-squared error may be minimized with respect to the $\{\lambda_n^2\}$ by minimizing the Lagrangian

$$\mathcal{L} = \left(\sum_{n=m+1}^{N}\lambda_n^2\right)^2 + 2\sum_{n=1}^{m}(\lambda_n^2)^2 - \mu\left(\frac{1}{N}\sum_{n=1}^{N}\lambda_n^2 - 1\right). \tag{84}$$

The resulting regression equations are

$$\begin{aligned}
n \le m : \frac{\partial\mathcal{L}}{\partial\lambda_n^2} &= 4\lambda_n^2 - \frac{\mu}{N} = 0 \\
: \lambda_n^2 &= \frac{\mu}{4N} \\
n > m : \frac{\partial\mathcal{L}}{\partial\lambda_n^2} &= 2\sum_{n=m+1}^{N}\lambda_n^2 - \frac{\mu}{N} = 0 \\
: \sum_{n=m+1}^{N}\lambda_n^2 &= \frac{2\mu}{4N}.
\end{aligned}$$

Invoke the constraint to solve for μ:

$$\frac{1}{N}\sum_{n=1}^{N}\lambda_n^2 = \frac{1}{N}\,\frac{m\mu}{4N} + \frac{1}{N}\,\frac{2\mu}{4N} = 1 \tag{85}$$

$$\mu = \frac{4N^2}{m+2}; \qquad \frac{\mu}{4N} = \frac{N}{m+2}.$$

The resulting minimum value of mean-squared error is

$$\begin{aligned}\mathrm{MSE}(m) &= \frac{r_0^2}{N^2}\left(\frac{2N}{m+2}\right)^2 + \frac{2r_0^2}{N^2}\,m\left(\frac{N}{m+2}\right)^2 \\ &= \frac{2r_0^2}{m+2}.\end{aligned} \tag{86}$$

This result establishes *the* lower bound on the mean-squared error of any low-rank estimator of the power in a band.

Known $\{\lambda_n^2\}_1^N$. The estimator

$$\hat{r}_0 = \beta\,\frac{1}{N}\sum_{n=1}^{m}\frac{1}{\lambda_n^2}\,\mathbf{y}^T P_n \mathbf{y} \tag{87}$$

is the obvious low-rank approximation of $\hat{r}_0$ when the $\{\lambda_n^2\}_1^N$ are known. The mean, variance, and mean-squared error of $\hat{r}_0$ are

$$E\,\hat{r}_0 = r_0\beta\,\frac{m}{N} \tag{88}$$

$$\mathrm{var}\,\hat{r}_0 = 2r_0^2\beta^2\,\frac{1}{N^2}\,m \tag{89}$$

$$\mathrm{MSE}(m) = r_0^2\left(1 - \frac{\beta m}{N}\right)^2 + 2r_0^2\beta^2\,\frac{1}{N^2}\,m. \tag{90}$$

The minimizing value of β is

$$\beta = \frac{N}{m+2}; \qquad 1 - \frac{\beta m}{N} = \frac{2}{m+2}, \tag{91}$$

and the corresponding minimum value of $\mathrm{MSE}(m)$ is

$$\begin{aligned}\mathrm{MSE}(m) &= r_0^2\left(\frac{2}{m+2}\right)^2 + 2r_0^2\left(\frac{1}{m+2}\right)^2 m \\ &= \frac{2r_0^2}{m+2}.\end{aligned} \tag{92}$$

This result matches our previous result, leading to the conclusion that the mean-squared error of any low-rank estimator of the power in a band is bounded as follows:

$$\mathrm{MSE}\Big(\int_{B(\phi)} \hat{S}_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi}\Big) \geq \frac{2}{m+2}\Big(\int_{B(\phi)} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi}\Big)^2. \tag{93}$$

3.6 Complex Demodulation and Frequency Sweeping

In Section 2.5 we argued that the power in the time series $\{x_t\}$ that is concentrated in the narrow spectral band $B(\phi)$ is twice the power in the complex demodulated time series $\{e^{-j\phi t}x_t\}$ that is concentrated in the baseband B. This property follows from the fact that the covariance sequence for the complex demodulated sequence is $\{e^{-j\phi t}r_t\}$ and the power spectrum is $S_{xx}(e^{j(\phi+\theta)})$.

Let's assume that the complex demodulated sequence is filtered by the ideal baseband filter $H(z)$, illustrated in Figure 3. The complex snapshot observed at the output of the filter is $\mathbf{y} = D(e^{-j\phi})\mathbf{x}$, and its covariance matrix is R_{yy}:

$$\mathbf{y} = D(e^{-j\phi})\mathbf{x} \tag{94}$$

$$\begin{aligned} R_{yy} = E\,\mathbf{y}\mathbf{y}^* &= \int_B S_{xx}(e^{j(\theta+\phi)})\Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi} \\ &= D(e^{-j\phi})\int_{B_+(\phi)} S_{xx}(e^{j\theta})\Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi}\,D(e^{j\phi}). \end{aligned} \tag{95}$$

We have used the identity $\Psi(e^{j(\theta-\phi)}) = D(e^{-j\phi})\Psi(e^{j\theta})$ to derive these two equivalent formulas for R_{yy}. The covariance matrix R_{yy} may be rewritten as

$$D(e^{j\phi})R_{yy}D(e^{-j\phi}) = \int_{B_+(\phi)} S_{xx}(e^{j\theta})\Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi}. \tag{96}$$

The covariance matrix on the right-hand side belongs to a complex snapshot whose spectrum is concentrated on the upper sideband $B_+(\phi)$. The covariance matrix on the left-hand side belongs to the modulated snapshot $D(e^{j\phi})\mathbf{y}$. The two snapshots are equivalent.

The power in the complex demodulated snapshot $\mathbf{y}$ is the power of the original signal that is concentrated in the upper sideband $B_+(\phi)$ and half the power of the original signal that is concentrated in the narrow spectral band $B(\phi)$:

$$\begin{aligned} E|y_t|^2 = \frac{1}{N}\operatorname{tr} R_{yy} &= \int_B S_{xx}(e^{j(\theta+\phi)})\,\frac{d\theta}{2\pi} \\ &= \int_{B_+(\phi)} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi} = \frac{1}{2}\int_{B(\phi)} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi}. \end{aligned} \tag{97}$$

If the power spectral density $S_{xx}(e^{j\theta})$ is slowly varying in the band $B(\phi)$, then the translated spectrum $S_{xx}(e^{j(\theta+\phi)})$ is slowly varying on the band B. This means that the power in the narrow band $B(\phi)$ is approximately

$$r_0(\phi) = \int_{B(\phi)} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi} = 2\beta S_{xx}(e^{j\phi}). \tag{98}$$

The covariance matrix R_{yy} is approximately

$$\begin{aligned} R_{yy} &\cong S_{xx}(e^{j\phi}) \int_B \Psi(e^{j\theta})\Psi^*(e^{j\theta}) \frac{d\theta}{2\pi} \\ &= S_{xx}(e^{j\phi})R = \frac{1}{2\beta} r_0(\phi) R \end{aligned} \tag{99}$$

where R is the baseband covariance matrix derived in Section 2.5:

$$R = \int_B \Psi(e^{j\theta})\Psi^*(e^{j\theta}) \frac{d\theta}{2\pi} \tag{100}$$

The trace of R is the time-bandwidth product $N\beta$ and $\frac{1}{N}\operatorname{tr} R_{yy}$ is half $r_0(\phi)$:

$$\frac{1}{N}\operatorname{tr}[R_{yy}] = \frac{1}{2} r_0(\phi). \tag{101}$$

Our ML results may be applied without change to the estimation of $r_0(\phi)$ in the model $\sqrt{2}\,\mathbf{y} : N[0, r_0(\phi)\frac{1}{\beta}R]$, with $\frac{1}{N}\operatorname{tr}\frac{1}{\beta}R = 1$. For example, when the eigenvectors of R are known but the eigenvalues are assumed unknown, the reduced rank estimator of $r_0(\phi)$ is

$$\hat{r}_0(\phi) = \frac{2}{N}||P\mathbf{y}||^2 = \frac{2}{N}\left|\left|PD(e^{-j\phi})\mathbf{x}\right|\right|^2 \tag{102}$$

$$P = \sum_{n=1}^{m} \mathbf{u}_n\mathbf{u}_n^T; \qquad \mathbf{u}_n : \text{eigenvector of } R\mathbf{u} = \lambda\mathbf{u}.$$

This estimator uses a complex demodulator to generate the snapshot $\mathbf{y}$ and the rank-one projections illustrated in Figure 9.

3.7 Reducing Rank and Removing the Bandpass Filter

Our most refined experiment for spectrum analysis is depicted in Figure 10(a). The complex demodulated signal $\{e^{-j\phi t}x_t\}$ is passed through the ideal baseband filter $H(z)$, the resulting snapshot $\mathbf{y}$ is projected through P onto the subspace spanned by the m dominant eigenvectors of R, and the power in that subspace is used to estimate the power in the narrowband $B(\phi)$. As illustrated in Figure 5, the projector P is very selective in frequency when the rank m is chosen to be the time bandwidth product $m = N\beta$. This means, for all practical purposes, that the ideal baseband filter $H(z)$ may be removed to produce the practical diagram of Figure 10(b). The diagram of Figure 10(b) may be redrawn as in Figure 10(c). In 10(c), a snapshot of data $\mathbf{x} = (x_0 x_1 \cdots x_{N-1})$ is complex demodulated with the demodulation matrix $D(e^{-j\phi})$ to produce the complex demodulated snapshot $\mathbf{y}$ that is then projected onto a subspace where the norm-squared is computed and scaled to estimate power.

3.8 Projection-Based Spectrum Analysis

The spectrum estimator

$$\hat{r}_0(\phi) = 2\beta \hat{S}_{xx}(e^{j\phi}) = \frac{2}{N} \left|\left|PD(e^{-j\phi})\mathbf{x}\right|\right|^2 \tag{103}$$

is nothing more nor less than the norm-squared of a complex demodulated snapshot that has been projected onto a narrowband subspace. A geometrical interpretation is presented in Figure 11. The complex demodulation matrix $D(e^{-j\phi})$ is a unitary, or rotation, matrix that simply rotates the measurement vector $\mathbf{x}$ into a position in complex Euclidean space where its projection onto the "baseband subspace $\langle U_m \rangle$" may be used to estimate the power in band $B(\phi)$. When this procedure is carred out for all ϕ, the power spectrum is mapped out. The result of equation (100) may be rewritten to make it look like Thomson's [8] multi-window spectrum estimator. Let's write the spectrum estimator as

$$\begin{aligned} \hat{S}_{xx}(e^{j\phi}) &= \frac{1}{N\beta} \left|\left| \sum_{n=1}^{m} \mathbf{u}_n \mathbf{u}_n^T D(e^{-j\phi})\mathbf{x} \right|\right|^2 \\ &= \frac{1}{N\beta} \sum_{n=1}^{m} \left|\mathbf{u}_n^T D(e^{-j\phi})\mathbf{x}\right|^2. \end{aligned} \tag{104}$$

In this form, the complex demodulated vector $D(e^{-j\phi})\mathbf{x}$ is correlated with the Slepian sequences $\{\mathbf{u}_n\}_1^m$, and the squared magnitudes are summed. This is illustrated in Figure 12(a). The inner product $\mathbf{u}_n^T D(e^{-j\phi})\mathbf{x}$ may also be written as follows:

$$\mathbf{u}_n^T D(e^{-j\phi})\mathbf{x} = \mathbf{x}^T W_n \Psi(e^{j\phi}), \tag{105}$$

where W_n is the diagonal window matrix constructed from the n^{th} Slepian sequence and $\Psi(e^{j\phi})$ is the Fourier vector:

$$\begin{aligned} W_n &= \mathrm{diag}\big(u_n(1)\, u_n(2) \cdots u_n(N)\big) \\ \mathbf{u}_n &= \big(u_n(1)\, u_n(2) \cdots u_n(N)\big)^T. \end{aligned} \tag{106}$$

The estimator

$$S_{xx}(e^{j\phi}) = \frac{1}{\beta}\sum_{n=1}^{m}\frac{1}{N}\left|\mathbf{x}^T W_n \Psi(e^{j\phi})\right|^2 \tag{107}$$

is a multi-window spectrum estimator wherein the sequence $\mathbf{x}$ is windowed with Slepian sequences, and the resulting sequence is used to compute the windowed periodogram $\frac{1}{N}\left|\mathbf{x}^T W_n \Psi(e^{j\phi})\right|^2$. Then m such periodograms are averaged, as illustrated in Figure 12(b).

One final interpretation—the multi-window, or projection-based, spectrum estimator is essentially a reduced rank, smoothed periodogram. To see this, consider the smoothed periodogram originally advocated by Daniell [5]:

$$\begin{aligned}
\hat{S}_{xx}(e^{j\phi}) &= \frac{2\pi}{|B(\phi)|}\int_{B(\phi)} \frac{1}{N}\left|\sum_{t=0}^{N-1} x_t e^{-j\theta t}\right|^2 \frac{d\theta}{2\pi} \\
&= \frac{2\pi}{2\beta\pi}\int_{B} \frac{1}{N}\left|\sum_{t=0}^{N-1} x_t e^{-j(\theta-\phi)t}\right|^2 \frac{d\theta}{2\pi} \\
&= \frac{1}{N\beta}\mathbf{x}^T D(e^{j\phi})\int_{B} \Psi(e^{j\theta})\Psi^*(e^{j\theta})\frac{d\theta}{2\pi} D(e^{-j\phi})\mathbf{x} \\
&= \frac{1}{N\beta}\mathbf{x}^T D(e^{j\phi}) R D(e^{-j\theta})\mathbf{x}.
\end{aligned} \tag{108}$$

This kind of quadratic form will be studied extensively in Sections 4.0 and 5.0. For the time being, we simply observe that, when the covariance matrix R is replaced by the approximation $R \cong \sum_{n=1}^{m}\mathbf{u}_n\mathbf{u}_n^T$, the Daniell smoothed periodogram becomes the projection-based, or multi-window, spectrum estimator

$$\begin{aligned}
\hat{S}_{xx}(e^{j\phi}) &\cong \frac{1}{N\beta}\sum_{n=1}^{m}\left|\mathbf{u}_n^T D(e^{-j\phi})\mathbf{x}\right|^2 \\
&= \frac{1}{N\beta}\left|\left|P D(e^{-j\phi})\mathbf{x}\right|\right|^2 \\
&= \frac{1}{N\beta}\sum_{n=1}^{m}\left|\Psi^*(e^{j\phi}) W_n \mathbf{x}\right|^2.
\end{aligned} \tag{109}$$

4.0 REPRESENTATION THEOREM FOR QUADRATIC ESTIMATORS

The covariance sequence $\{r_t\}$ and the power in a narrow spectral band $B(\phi)$ are the two most elementary linear functionals of a power spectrum that are encountered in the study of WSS time series. The power in a narrow spectral band, namely

$$r_0(\phi) = \int_{B(\phi)} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi}. \tag{110}$$

characterizes the local power of the spectrum. The trigonometric functionals

$$r_t = \int_{-\pi}^{\pi} S_{xx}(e^{j\theta}) e^{jt\theta} \frac{d\theta}{2\pi} \tag{111}$$

characterize the entire power spectrum $S_{xx}(e^{j\theta})$ via

$$S_{xx}(e^{j\theta}) = \sum_t r_t e^{-jt\theta}. \tag{112}$$

Linear Functionals of the Power Spectrum. The experiment illustrated in Figure 13 shows how other quadratic forms may arise naturally. The sequences $\{u_t\}$ and $\{v_t\}$ are the outputs of two filters, $F(z)$ and $G(z)$, driven by the common input sequence $\{x_t\}$. The zero-lag cross-covariance between the outputs is

$$E\, u_t v_t^* = \int_{-\pi}^{\pi} F(e^{j\theta}) S_{xx}(e^{j\theta}) G^*(e^{j\theta}) \frac{d\theta}{2\pi}. \tag{113}$$

The quantity $u_t v_t^*$ is quadratic in the data $\{x_t\}$, and the expectation of this quadratic form is a linear functional of the power spectrum $S_{xx}(e^{j\theta})$. In fact, the expected value of every quadratic form in $\{x_t\}$ is a linear functional of the power spectrum $S_{xx}(e^{j\theta})$.

If we take

$$F(e^{j\theta}) = G(e^{j\theta}) = H_B(e^{j\theta}) \tag{114}$$

(where $H_B(e^{j\theta})$ is the ideal bandpass filter of Figure 1), then the outputs u_t and v_t are

$$u_t = v_t = y_t. \tag{115}$$

The sequence $\{y_t\}$ is that part of $\{x_t\}$ lying in the band $B(\phi)$. Equation (113) is then

$$E(|y_t|^2) = \int_{-\pi}^{\pi} |H_B(e^{j\theta})|^2 S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi}. \tag{116}$$

which is the same as equation (6).

If we were to let

$$G(e^{j\theta}) = 1, \tag{117}$$

then $y_t = x_t$ and the right-hand side of equation (113) becomes

$$\int_{-\pi}^{\pi} F(e^{j\theta}) S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi},$$

which is a "general" linear functional of the power spectrum $S_{xx}(e^{j\theta})$. Every such linear functional is the expected value of a quantity which is quadratic in the signal $\{x_t\}$.

Finite Quadratic Forms. Every quadratic function of the *finite* data vector $\mathbf{x} = [x_0\ x_1 \cdots x_{N-1}]^T$ has expectation

$$E[\mathbf{x}^* Q\mathbf{x}] = \int_{-\pi}^{\pi} F(e^{j\theta}) S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi}, \tag{118}$$

where

$$F(e^{j\theta}) = \sum_{k=0}^{N-1} \sum_{m=0}^{N-1} q_{km} e^{-j(k-m)\theta}. \tag{119}$$

(This is a trigonometric polynomial.) Every such quadratic form delivers an unbiased estimate of a linear functional of $S(e^{j\theta})$. The family of all such quadratic forms provides information about the power spectrum, but not enough to completely characterize it. Since

$$E[\mathbf{x}^* Q\mathbf{x}] = \sum_{k=0}^{N-1} \sum_{m=0}^{N-1} q_{km} r_{k-m}, \tag{120}$$

we see that only those values of r_k in the range $-N < k < N$ are involved. Estimates of r_t outside this range necessarily depend on some "model" or functional form for $S(e^{j\theta})$ or involve extension principles like maximum entropy (which forces a model indirectly).

If one does not want to assume a model, then it becomes necessary to admit that we can know something about $S_{xx}(e^{j\theta})$ but that we cannot know it completely. Each linear functional of $S_{xx}(e^{j\theta})$ forces it into a hyperplane in function space. Knowing the values of n linear functionals forces $S_{xx}(e^{j\theta})$ into the intersection of n hyperplanes. This is still an infinite dimensional space.

4.1 Estimator Properties

Assuming that we have agreed to consider only quadratic estimators, we must still decide which quadratic form to use and we must decide how compelling the results of Sections 2.0 and 3.0 are. As a tool for eliminating undesirable choices, let us list some properties which power spectrum estimators of the form

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \mathbf{x}^* Q(\theta)\mathbf{x} \tag{121}$$

should enjoy.

Positivity. The power spectrum should be nonnegative real,

$$\hat{S}(e^{j\theta}, \mathbf{x}) \geq 0 \tag{122}$$

This condition will hold if the matrix $Q(\theta)$ is positive semidefinite. We will use the notation

$$Q(\theta) \geq 0 \tag{123}$$

for this.

The rest of the properties we shall list all involve linear transformations on the data vector $\mathbf{x}$.

Amplitude Scaling. If the signal $\mathbf{x}$ is multiplied by the scalar μ, then the power spectrum should be multiplied by $|\mu|^2$. Therefore,

$$\hat{S}(e^{j\theta}, \mu\mathbf{x}) = |\mu|^2 \hat{S}(e^{j\theta}, \mathbf{x}). \tag{124}$$

This property holds for any quadratic estimator.

Modulation Invariance. Suppose we construct a signal

$$y_t = e^{jt\phi} x_t. \tag{125}$$

This is called *modulation* and shifts the signal in frequency by the angle ϕ. To see this, consider the autocorrelation sequence

$$\begin{aligned} E[y_{t+m} y_m^*] &= e^{j(t+m-m)\phi} E\left[x_{t+m} x_m^*\right] \\ &= e^{jt\phi} r_t. \end{aligned} \tag{126}$$

Taking Fourier transforms yields the frequency domain equivalent,

$$S_{yy}(e^{j\theta}) = S_{xx}(e^{j(\theta-\phi)}). \tag{127}$$

For finite data, the linear transformation in equation (125) has matrix representation

$$\mathbf{y} = D(e^{j\phi})\mathbf{x}, \tag{128}$$

where

$$D(e^{j\phi}) = \mathrm{diag}[1, e^{j\phi}, e^{j2\phi}, \ldots, e^{j(N-1)\phi}]. \tag{129}$$

The estimator property consistent with equation (127) is

$$\hat{S}(e^{j\theta}, D(e^{j\phi})\mathbf{x}) = \hat{S}(e^{j(\theta-\phi)}, \mathbf{x}). \tag{130}$$

We shall call an estimator with this property *modulation invariant.*

If $\hat{S}_0(e^{j\theta}, \mathbf{x})$ is an estimator which does not have this property, then we can construct a one-parameter family of estimators

$$\hat{S}_\phi(e^{j\theta}, \mathbf{x}) = \hat{S}_0(e^{j(\theta+\phi)}, D(e^{j\phi})\mathbf{x}). \tag{131}$$

There is no *a priori* reason to prefer one member of the family over another. The average

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \int_{-\pi}^{\pi} \hat{S}_\phi(e^{j\theta}, \mathbf{x}) \, \frac{d\phi}{2\pi} \tag{132}$$

would be modulation invariant. Modulation invariance is not unique to quadratic estimators. Most common autoregressive estimators have the property, even when they are constructed from the normal equations of linear prediction.

J-**Symmetry.** Suppose we construct the time-reversal of the signal $\{x_t\}$:

$$y_t = x_{-t}. \tag{133}$$

Then

$$E\, y_{t+k} y_t^* = r_{-k} = r_k^* \tag{134}$$

and

$$S_{yy}(e^{j\theta}) = S_{xx}(e^{-j\theta}). \tag{135}$$

This produces something different only if the signal x is not real-valued. As estimator version of this is

$$\hat{S}(e^{j\theta}, J\mathbf{x}) = \hat{S}(e^{-j\theta}, \mathbf{x}). \tag{136}$$

We will call such as estimator *J-symmetric*. The matrix

$$J = \begin{bmatrix} \mathcal{O} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \mathcal{O} \end{bmatrix} \tag{137}$$

reverses the components of any vector to which it is applied.

Band Limiting. Let y_t be the signal generated in the ideal experiment of Figure 1. This is that part of the signal $\{x_t\}$ which lies in the band $B(\phi)$. Then

$$\begin{aligned} S_{yy}(e^{j\theta}) &= \left|H_B(e^{j\theta})\right|^2 S_{xx}(e^{j\theta}) \\ &= \begin{cases} S_{xx}(e^{j\theta}), & \theta \in B(\phi) \\ 0, & \theta \notin B(\phi). \end{cases} \end{aligned} \tag{138}$$

It is (unfortunately) impossible to generate any value of y_t from a finite record of the signal $\{x_t\}$. We would like to write

$$y = P_B x, \tag{139}$$

where P_B is some matrix which projects the data $\{x_t\}$ onto the "subspace of signals lying in the band $B(\phi)$." Equation (139) is exact only in the infinite dimensional case. If it held in the finite dimensional case, we could demand a property of the form

$$\hat{S}(e^{j\theta}, P_B\mathbf{x}) = H_B(e^{j\theta})\hat{S}(e^{j\theta}, \mathbf{x}). \tag{137}$$

Although this property is unattainable with finite data, our estimator should approximate it. To the extent that we can construct some family of "bandpass filtering" projection matrices for which this holds down to a certain bandwidth, then we can say that we have "resolution" to that bandwidth. Without it, we can construct estimators which enjoy all the properties previously mentioned but which are trivial. An example of such as estimator would be

$$\hat{S}(e^{j\theta}, \mathbf{x}) = ||\mathbf{x}||^2.$$

We shall not demand as yet that our estimator enjoy all the properties mentioned in the previous section. To begin. we require a quadratic estimator which is positive and modulation invariant. With this minimum requirement, we prove the following representation theorem.

Representation Theorem: Every non-negative, quadratic, modulation invariant power spectrum estimator has the form

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \left|\left|V^* D(e^{-j\theta})\mathbf{x}\right|\right|^2. \tag{141}$$

where V is an $m \times N$ complex matrix having rank m and $D(e^{j\theta})$ is defined in equation (129).

Proof: If the estimate is quadratic, then

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \mathbf{x}^* Q(\theta)\mathbf{x}, \tag{142}$$

and if it is also nonnegative then

$$Q(\theta) \geq 0$$

for each θ. In particular, $Q(0) > 0$ and can therefore be factored as

$$Q(0) = V^* V, \tag{143}$$

where V is $m \times N$ and

$$m = \operatorname{rank} Q(0).$$

If the estimator is modulation invariant, then

$$\left[D(e^{j\phi})\mathbf{x}\right]^* Q(\theta)\left[D(e^{j\phi})\mathbf{x}\right] = \mathbf{x}^* Q(\theta - \phi)\mathbf{x}$$

for all $\mathbf{x}$, θ, ϕ. This implies that

$$Q(\theta - \phi) = D^*(e^{j\phi})Q(\theta)D(e^{j\phi})$$

for all θ, ϕ. Take $\theta = 0$ and reverse the sign of ϕ to get

$$\begin{aligned} Q(\phi) &= D(e^{-j\phi})Q(0)D(e^{-j\phi}) \\ &= D(e^{j\phi})V^* V D(e^{-j\phi}). \end{aligned} \tag{144}$$

Combining equations (144) and (142) yields

$$\begin{aligned} \hat{S}(e^{j\theta}, \mathbf{x}) &= \mathbf{x}^* D(e^{j\theta})V^* V D(e^{-j\theta})\mathbf{x} \\ &= \left|\left|V D(e^{-j\theta})\mathbf{x}\right|\right|^2. \end{aligned} \tag{145}$$

∎

4.2 Multiple Window Interpretation

The estimator of equation (141) can be understood as the composition of three operations. Multiplying the data by $D(e^{-j\theta})$ is *demodulation*, or shifting in frequency, so that what was at angle θ is now at zero. Operating on the result by V is akin to passing the demodulated data through a bank of m lowpass filters. The average energy of the resulting outputs is estimated by taking the norm squared. This yields an estimate of the power in the vicinity of θ.

We can put the computation of $\hat{S}(e^{j\theta}, \mathbf{x})$ in another way which allows for the use of an FFT operation to provide samples. Write

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \left|\left|VD(e^{-j\theta})\mathbf{x}\right|\right|^2 = \sum_{\mu=1}^{m} \left| \sum_{k=0}^{N-1} e^{-jk\theta} [V_{\mu k} x_k] \right|^2. \tag{146}$$

In this form, the estimator is seen to be a generalization of the classical periodogram, which is

$$\hat{S}_1(e^{j\theta}, \mathbf{x}) = \frac{1}{N} \left| \sum_{k=0}^{N-1} x_k e^{-jk\theta} \right|^2. \tag{147}$$

In this case, $m = 1$ and

$$V = \frac{1}{\sqrt{N}} \mathbf{1}^T, \tag{148}$$

where

$$\mathbf{1}^T \triangleq [1, 1, \dots, 1]. \tag{149}$$

If the data is first "windowed" by multiplying pointwise with a window sequence, then we will have the periodogram with window

$$\hat{S}_w(e^{j\theta}, \mathbf{x}) = \frac{1}{N} \left| \sum_{k=0}^{N-1} x_k w_k e^{-jk\theta} \right|^2. \tag{150}$$

The general quadratic, positive, modulation invariant estimator in equation (146) is simply an average of m windowed periodograms. For this reason, it is called a ***multiple window periodogram***. Since it amounts to the sum of m periodograms, one can efficiently obtain samples of $\hat{S}(e^{j\theta}, \mathbf{x})$ with spacing $2\pi/N$ by employing m FFT operations.

4.3 J-Symmetry

We shall call any estimator of the form in equation (141) a ***natural estimator***. It is completely characterized by the $m \times N$ matrix V. This matrix is, however, not uniquely determined by the estimator since multiplication on the left by a unitary matrix U will produce $V' = UV$ and leave the estimate unchanged. If U is $m \times m$ and

$$U^* U = I, \tag{151}$$

then

$$||VD(e^{-j\theta})\mathbf{x}||^2 = ||UVD(e^{-j\theta})\mathbf{x}||^2. \tag{152}$$

In general, the matrices V and V' generate the same natural estimator if and only if there exists a unitary matrix U for which

$$V' = UV. \tag{153}$$

This observation is relevant to the following question.

Under what conditions is a natural estimator J-symmetric—i.e. it satisfies the identity (136)? If we assume that both equations (136) and (141) hold, then

$$||VD(e^{-j\theta})J\mathbf{x}||^2 = ||VD(e^{j\theta})\mathbf{x}||^2 \text{ for all } \theta, \mathbf{x}. \tag{154}$$

We make the observation that the matrices defined in equations (129) and (137) satisfy

$$D(e^{-j\theta})J = e^{-j(N-1)\theta} JD(e^{j\theta}). \tag{155}$$

Thus the estimator is J symmetric if and only if

$$||VJD(e^{j\theta})\mathbf{x}||^2 = ||VD(e^{j\theta})\mathbf{x}||^2 \text{ for all } \theta, \mathbf{x}. \tag{156}$$

In other words, $V' = VJ$ produces the same estimator as does V. It must therefore have the form $V' = UV$ for some unitary matrix U.

In summary, a natural estimator is also J-symmetric if and only if

$$VJ = UV \tag{157}$$

for some unitary matrix U.

4.4 The Mean of the Estimator

The J-symmetry and modulation invariance properties provide gross information about the form (141) of natural estimators and the matrix V (equation (157)). We do not as yet have a tool for understanding the resolution or frequency discrimination properties of the estimator. Intuitively, we know that the rows of V can be thought of as FIR lowpass filter responses and as window functions. The resolution would then be somehow related to the bandwidth of the frequency response functions of these m filters. But why use m competing lowpass filters instead of one "good" one? In order to come to an understanding of these phenomena, we are led to evaluate the mean and variance of natural estimators. The mean will depend on $S(e^{j\theta})$ and the matrix V. The same will be true of the variance provided we make the assumption that the signal x is Gaussian so that we can characterize fourth moments in terms of $S(e^{j\theta})$.

Since the estimator $\hat{S}(e^{j\theta}, \mathbf{x})$ is quadratic in the data vector $\mathbf{x}$, its expectation will be a linear functional of $S(e^{j\theta})$. We shall construct two expressions for this.

Let R_{xx} be the covariance of the data vector $\mathbf{x}$:

$$R_{xx} = E[xx^*] = \begin{bmatrix} r_0 & r_{-1} & \cdots & r_{-N+1} \\ r_1 & & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_{-1} \\ r_{N-1} & \cdots & r_1 & r_0 \end{bmatrix}. \tag{158}$$

Using the notation of Section 3.0, we may use the Fourier representation (43) to obtain a representation for the matrix R_{xx}:

$$R_{xx} = \int_{-\pi}^{\pi} S_{xx}(e^{j\theta})\Psi(e^{j\theta})\Psi^*(e^{j\theta})\,\frac{d\theta}{2\pi}. \tag{159}$$

Now let us compute the mean of $\hat{S}(e^{j\theta}, \mathbf{x})$:

$$\begin{aligned} E\big[\hat{S}(e^{j\theta}, \mathbf{x})\big] &= E\big[\mathbf{x}^* D(e^{j\theta})V^*VD(e^{-j\theta})\mathbf{x}\big] \\ &= \mathrm{Tr}\big[D(e^{j\theta})V^*VD(e^{-j\theta})R_{xx}\big] \\ &= \mathrm{Tr}\big[VD(e^{-j\theta})R_{xx}D(e^{j\theta})V^*\big]. \end{aligned} \tag{160}$$

This is the first expression for the mean, and it uses the covariance matrix R_{xx}. This can be reduced to a linear functional of $S(e^{j\theta})$ by using the representation (159) for R_{xx}. The result is

$$E\big[\hat{S}(e^{j\theta}, \mathbf{x})\big] = \int_{-\pi}^{\pi} S(e^{j\phi})\big|\big|V\Psi(e^{j(\theta-\phi)})\big|\big|^2\,\frac{d\phi}{2\pi}. \tag{161}$$

This bears comment. Let

$$W(z) = V\Psi(z) = \begin{bmatrix} W_1(z) \\ \vdots \\ W_m(z) \end{bmatrix}. \tag{162}$$

Then $W_\mu(e^{j\theta})$ is the frequency response of the μ^{th} row of V, which we think of as a lowpass filter. The mean of the estimate is the convolution of the true spectrum with the positive real function

$$\big|\big|W(e^{j\theta})\big|\big|^2 = \sum_{\mu=1}^{m} \big|W_\mu(e^{j\theta})\big|^2. \tag{163}$$

This is the sum of the magnitude-squared responses of the m lowpass filters. If these filters have small bandwidth, then $E[\hat{S}(e^{j\theta}, \mathbf{x})]$ will be a weighted average of the values of the true spectrum on an interval of the same width centered at θ. This is the means by which we can judge the resolution of the estimator.

4.5 Variance and Mean-Squared Error of the Estimator

We must assume that the data vector x is real valued and Gaussian with mean vector $\mathbf{0}$ and covariance matrix R_{xx}. If

$$Q = A + jB \tag{164}$$

is Hermitian, then the random variable

$$y = \mathbf{x}^T Q \mathbf{x} = \mathbf{x}^T A \mathbf{x} \tag{165}$$

has mean and variance

$$E[\mathbf{y}] = \mathrm{Tr}[QR_{xx}] = \mathrm{Tr}[AR_{xx}] \tag{166}$$

$$\mathrm{var}(\mathbf{y}) = 2\mathrm{Tr}\left[(AR_{xx})^2\right]. \tag{167}$$

Our estimator has this form, with

$$Q(\theta) = D(e^{j\theta})V^{\star}VD(e^{-j\theta}) \tag{168}$$

and

$$A(\theta) = \frac{1}{2}\left[Q(\theta) + Q(-\theta)\right]. \tag{169}$$

Therefore,

$$\mathrm{var}\,\hat{S}(e^{j\theta}, \mathbf{x}) = \frac{1}{2}\mathrm{Tr}\left[(Q(\theta) + Q(-\theta))R_{xx}\right]^2. \tag{170}$$

If we expand the square in this expression, we will arrive at four terms. It turns out that two of these have the same trace, and the remaining two have the same trace. Define

$$M(\theta) = VD(e^{-j\theta})R_{xx}D(e^{j\theta})V^{\star} \tag{171}$$

$$N(\theta) = VD(e^{j\theta})R_{xx}D(e^{-j\theta})V^{\star}. \tag{172}$$

Using the trace identity

$$\mathrm{Tr}(FG) = \mathrm{Tr}(GF) \tag{173}$$

liberally, we see that

$$\mathrm{Tr}\left[Q(\theta)R_{xx}Q(\theta)R_{xx}\right] = \mathrm{Tr}\left[M^2(\theta)\right] \tag{174}$$

$$\mathrm{Tr}\left[Q(\theta)R_{xx}Q(-\theta)R_{xx}\right] = \mathrm{Tr}\left[N(-\theta)N(\theta)\right]$$

$$\begin{aligned}\mathrm{Tr}\left[Q(-\theta)R_{xx}Q(\theta)R_{xx}\right] &= \mathrm{Tr}\left[N(\theta)N(-\theta)\right]\\ &= \mathrm{Tr}\left[N(-\theta)N(\theta)\right]\end{aligned}$$

$$\mathrm{Tr}\left[Q(-\theta)R_{xx}Q(-\theta)R_{xx}\right] = \mathrm{Tr}\left[M^2(-\theta)\right].$$

The middle two terms have the same trace. In addition, $M^2(\theta)$ and $M^2(-\theta)$ differ only in imaginary part. Since they are both Hermitian, they have real trace, and therefore they have the same trace. Combining all this yields the variance expression

$$\begin{aligned}\mathrm{var}\,\hat{S}(e^{j\theta},\mathbf{x}) &= \mathrm{Tr}\left[M^2(\theta)\right] + \mathrm{Tr}\left[N(\theta)N(-\theta)\right]\\ &= \mathrm{Tr}\left[M(\theta)M^*(\theta)\right] + \mathrm{Tr}\left[N(\theta)N^*(\theta)\right]\\ &= \sum_{\mu=1}^{M}\sum_{k=1}^{M}\left[|M_{\mu k}(\theta)|^2 + |N_{\mu k}(\theta)|^2\right].\end{aligned} \tag{175}$$

We can put this in terms of $S(e^{j\theta})$ by using the representation (159) for the matrix R_{xx} in the definitions for $M(\theta)$ and $N(\theta)$. This yields

$$M_{\mu k}(\theta) = \int_{-\pi}^{\pi} S_{xx}(e^{j\theta})W_\mu(e^{j(\theta-\phi)})W_k^*(e^{j(\theta-\phi)})\,\frac{d\phi}{2\pi} \tag{176}$$

$$N_{\mu k}(\theta) = \int_{-\pi}^{\pi} S_{xx}(\phi)W_\mu(e^{-j(\theta+\phi)})W_k^*(e^{j(\theta-\phi)})\,\frac{d\phi}{2\pi}. \tag{177}$$

Of the two variance terms, the one involving $M(\theta)$ will likely be much greater than the one involving $N(\theta)$ because of the misalignment of the window response functions in equation (177). From equations (160) and (175), we can write

$$E\left[\hat{S}(e^{j\theta},\mathbf{x})\right] = \mathrm{Tr}\left[M(\theta)\right] \tag{178}$$

$$\mathrm{var}\left[\hat{S}(e^{j\theta},\mathbf{x})\right] \geq \mathrm{Tr}\left[M^2(\theta)\right]. \tag{179}$$

The inequality is off by the term involving $N(\theta)$. These can be used to construct a lower bound on mean-squared error:

$$
\begin{aligned}
\mathrm{MSE}(\theta) &= E\big[S_{xx}(e^{j\theta}) - \hat{S}(e^{j\theta},\mathbf{x})\big]^2 \\
&= \big[S_{xx}(e^{j\theta}) - E\hat{S}(e^{j\theta},\mathbf{x})\big]^2 + \mathrm{var}\big[\hat{S}(e^{j\theta},\mathbf{x})\big] \qquad (180)\\
&\geq \big[S_{xx}(e^{j\theta}) - \mathrm{Tr}M(\theta)\big]^2 + \mathrm{Tr}M^2(\theta).
\end{aligned}
$$

Let the eigenvalues of $M(\theta)$ be $\{\mu_1, \ldots, \mu_m\}$. Then

$$
\mathrm{MSE}(\theta) \geq \left[S_{xx}(e^{j\theta}) - \sum_{k=1}^{M} \mu_k\right]^2 + \sum_{k=1}^{M} \mu_k^2. \qquad (181)
$$

The minimum of this expression, as a function of the eigenvalues, occurs when

$$
\mu_k = \frac{S_{xx}(e^{j\theta})}{m+1} \text{ for each } k. \qquad (182)
$$

Placing the minimum in the right-hand side of (171) produces the inequality

$$
\mathrm{MSE}(\theta) \geq \frac{S_{xx}^2(e^{j\theta})}{m+1}. \qquad (183)
$$

Since $M(\theta)$ is Hermitian, it can have identical eigenvalues only if it is a scalar multiple of the identity—i.e.

$$
M(\theta) = VD^*(\theta)R_{xx}D(\theta)V^* = \frac{S_{xx}(e^{j\theta})}{m+1} I. \qquad (184)
$$

The lower bound (183) provides a reason for using more than one window. (m is the number of windows and the number of rows in V.)

5.0 SPECIAL QUADRATIC ESTIMATORS

The spectral estimator

$$\hat{S}(e^{j\theta}, \mathbf{x}) = ||VD(e^{-j\theta})\mathbf{x}||^2 \tag{185}$$

of equation (141) composes three operations: demodulation, lowpass filtering, and energy measurement. It is interesting to compare this chain of operations with a superheterodyne radio receiver. The heart of such a receiver is a high quality bandpass filter (the IF amplifier) with a fixed center frequency and bandwidth. The rest of the receiver merely accomodates this filter. The incoming broadband signal is frequency shifted so that the desired band is moved into the IF band. On the other side, the output of the IF filter is "detected." The sensitivity of the receiver depends mostly on the quality of the IF filter. For our quadratic spectrum estimator, the quality (variance and resolution) depends entirely on the choice of lowpass filters represented by the rows of the matrix V.

In this section, we will present some likely choices of V in a somewhat developmental order beginning with rank-one estimators. Since several classical estimators have the form of equation (141), we will display them when appropriate.

5.1 Rank-One Estimators

For a rank-one estimator, there is but one lowpass filter, and the matrix V is a row vector. The estimator has the form of equation (150). That is.

$$\hat{S}(\theta, x) = \frac{1}{N}\left|\sum_{k=0}^{N-1} x_k w_k e^{-jk\theta}\right|^2, \tag{186}$$

and

$$V_{1k} = \frac{w_k}{\sqrt{N}}. \tag{187}$$

There is a choice of interpretations.

As before, we can think of demodulation, followed by lowpass filtering. In this interpretation, the matrix V represents a lowpass filter pulse response sequence, and we are using only one value of the lowpassed output sequence to estimate power.

The more common point of view is that we first "window" that data x_k by multiplying by w_k then demodulate and perform a time average. The shape of the window function

$$W(e^{j\theta}) = \sum_{k=0}^{N-1} w_k e^{-jk\theta} \tag{188}$$

will affect both the resolution (bias) and variance of the estimate. The windowing operation may be considered a form of *frequency averaging* applied to the data. Using the inverse transform

$$w_k = \int_{-\pi}^{\pi} W(e^{j\theta}) e^{jk\theta} \frac{d\theta}{2\pi}, \tag{189}$$

we see that the vector whose elements are $w_k x_k$ can be written

$$\int_{-\pi}^{\pi} \left[D(e^{j\theta})\mathbf{x}\right] W(e^{j\theta}) \frac{d\theta}{2\pi}. \tag{190}$$

This is a weighted average of the frequency translates of the data, $D(e^{j\theta})\mathbf{x}$. Therefore, to the extent that the window function $W(e^{j\theta})$ is dispersed, frequency resolution is degraded.

5.2 Low-Rank Estimators and the Time Bandwidth Product

The virtue of using more than one lowpass filter—i.e. of using a rank m estimator with $m > 1$—lies in the potential of decreased estimator variance. Up to a point, the mean-squared error should be inversely proportional to m. For given resolution requirements, however, there is a limiting value of m beyond which no improvement can be expected. This limit is a "time bandwidth product." In this section, we will put forth an argument for this assertion.

The choice of normalized bandwidth β, with

$$0 < \beta < 1,$$

is somewhat arbitrary. It is a rough measure of the resolution in frequency we expect from our spectrum estimator. If we assume that the power spectrum is essentially constant on intervals of length $2\pi\beta$, then an estimate of the power in the band $|\theta - \theta_0| < \beta\pi$ serves as an estimate of $\beta S_{xx}(e^{j\theta_0})$. In other words, we get an estimate of $S_{xx}(e^{j\theta_0})$ by estimating the power in a band about θ_0. This estimate will improve as we increase β, but only at the expense of resolution. This tradeoff is inescapable, and the sum of ill effects diminishes only with increasing data length N. For modulation-invariant estimators, the problem can be reduced to the estimate of power in the low-frequency band

$$|\theta| < \beta\pi, \tag{191}$$

(with $\theta_0 = 0$). This is the passband of the bank of lowpass filters described by the rows of the matrix V. For convenience, call this band *baseband*.

For the purpose of estimating *baseband power*

$$\int_{-\beta\pi}^{\beta\pi} S_{xx}(e^{j\theta})\,\frac{d\theta}{2\pi} \approx \beta S_{xx}(e^{j0}), \tag{192}$$

we would like to isolate the baseband component of the signal. But of course, we have only a finite time window's worth of data. We are caught in a classic dilemma—trying to isolate a signal (or a signal component) in both time and frequency. The more concentrated a signal becomes in one domain, the more dispersed it must be in the other. Although we cannot do this simultaneous isolation exactly, we must do it approximately if we are to do spectrum estimation.

Both operations—isolation in time (windowing) and isolation in frequency (lowpass filtering)—are projection operations. For the moment, let x be the entire WSS signal. Let P_T be the projection which zeros out all values of the signal except those in the data window:

$$(P_T x)_k = \begin{cases} x_k, & 0 \le k \le N-1 \\ 0, & \text{otherwise.} \end{cases} \tag{193}$$

Let P_Ω be the ideal baseband filter:

$$(P_\Omega x)_k = \sum_{l=-\infty}^{\infty} h_l x_{k-l}, \tag{194}$$

where

$$\sum_{k=-\infty}^{\infty} h_k e^{-jk\theta} = \begin{cases} 1, & |\theta| < \beta\pi \\ 0, & \text{otherwise.} \end{cases} \tag{195}$$

Each of these operators is an orthogonal projection:

$$P = P^2 = P^*. \tag{196}$$

But they do *not* commute. (The baseband component of the time windowed signal is not the time restricted baseband component.) If they did commute, then the product

$$P = P_T P_\Omega \tag{197}$$

would also be a projection. This projection would simultaneously isolate a signal component in both time and frequency.

In certain respects, however, the products are approximately equal:

$$P_T P_\Omega \approx P_\Omega P_T, \tag{198}$$

and the product is close to a projection having rank $N\beta$, the time-bandwidth product. This assertion can be supported in various ways, each of which corresponds to known multiple window spectrum estimators. We will sketch some of these in the following sections. For the moment, let us consider the consequences—on estimator variance—of the assumptions that equations (197) and (198) hold.

Suppose that

$$P = P_T P_\Omega = P_\Omega P_T, \tag{199}$$

is a rank $m = N\beta$ projection, and assume that the WSS signal x is Gaussian. Then Px is Gaussian and is concentrated on a subspace of dimension m, even though there may be N values of the vector which differ from zero. At most m scalar functions of Px can be *statistically* independent. In particular, at most m independent estimates of baseband power can be obtained, and therefore we can expect an improvement of $1/m$ in estimator variance when these are averaged. But we cannot expect any more than this.

If we actually had our hypothetical projection P, how should we estimate $S_{xx}(e^{j0})$? If $P = P_T P_\Omega$, then $Px = P_T y$, where y is the WSS baseband component of the WSS signal x. Therefore,

$$E[y_k^2] = \int_{-\beta\pi}^{\beta\pi} S_{xx}(e^{j\theta}) \frac{d\theta}{2\pi} \approx \beta S_{xx}(e^{j0}) \tag{200}$$

(assuming that the spectrum is fairly constant on intervals of width $2\beta\pi$). Now, using the definition of P_T,

$$\frac{1}{N} \|P_T \mathbf{y}\|^2 = \frac{1}{N} \sum_{k=0}^{N-1} y_k^2. \tag{201}$$

If we use this as an estimate of the variance of y_k and combine it with equation (200), we get an estimate for $S_{xx}(e^{j0})$:

$$\hat{S}(e^{j0}) = \frac{1}{N\beta} \|P_T \mathbf{y}\|^2 = \frac{1}{N\beta} \|P\mathbf{x}\|^2. \tag{202}$$

This estimate is obviously quadratic in the data and has rank $m = N\beta$, the time bandwidth product.

We have been using the notation x for the entire signal

$$x = \{x_k : -\infty < k < \infty\}.$$

The data vector which appears in the estimator (185) is therefore

$$\begin{bmatrix} x_0 \\ \vdots \\ x_{N-1} \end{bmatrix} = F^T x, \tag{203}$$

where

$$F^T = [\; \bigcirc \;\; I \; \bigcirc \;] \tag{194}$$

is an $N \times \infty$ matrix. The connection between the estimates (185) and (202) would be

$$\hat{S}(e^{j0}, F^T \mathbf{x}) = \|VF^T\mathbf{x}\|^2 = \frac{1}{N\beta}\|P\mathbf{x}\|^2. \tag{205}$$

This would determine V up to multiplication on the left by an orthogonal matrix, via

$$FV^TVF^T = \frac{1}{N\beta} P$$

and therefore (since $F^TF = I$)

$$V^TV = \frac{1}{N\beta} F^TPF. \tag{206}$$

This formula of course involves a projection P which does not exist. It is useful only for guessing what should actually be done. Let us now examine some approaches.

5.3 Time Division Multiple Windows

In [6], Welch proposed a quadratic multiple window estimator in which the matrix V would have the form

$$V = \begin{bmatrix} \cdots\cdots & & & & \\ & \cdots\cdots & & & \\ & & \cdots\cdots & & \\ & & & \cdots\cdots & \\ & & & & \cdots\cdots \end{bmatrix}. \tag{207}$$

The first row contains a symmetric window of length much less than N, followed by all zeros. The remaining rows are shifts of the first row, by equal amounts, so that the last row is right-justified. This matrix satisfies the J-symmetry condition

$$VJ = JV.$$

with V simply an $m \times m$ version of J. Let us give an interpretation of this estimator.

As before, let $y = P_\Omega x$ be the baseband component of the WSS signal x. Since y is bandlimited, it is completely determined by the subsequence of samples spaced $1/\beta$ apart (assuming that this is an integer). Loosely speaking, then, from every large set of N consecutive values of y, only $N\beta$ are useful. Moreover, if $S(e^{j\theta})$ is constant on baseband, then the subsamples of y (at spacing $1/\beta$) are indeed uncorrelated. Furthermore, the covariance matrix of

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix} = F^T y \tag{208}$$

should have $m = N\beta$ essentially equal dominant eigenvalues, with the other $N - m = (1-\beta)N$ very close to zero. This suggests that the baseband power estimate

$$\frac{1}{m} \sum_{k=0}^{m-1} y_{k/\beta}^2 \tag{209}$$

would have about the same variance as an estimate which used all N values of y but one-m^{th} the variance of an estimate which used only one value of y.

The Welch estimator approximates the scenario we have just described. The first row of V contains a properly normalized lowpass filter unit pulse response. Because of the successive shifts, the vector

$$V F^T x$$

represents m samples of the output of an FIR filter, equally spaced with spacing

$$\frac{1}{\beta} = \frac{N}{m},$$

so that

$$\beta \| V F^T x \|^2 = \frac{1}{m} \sum_{k=0}^{m-1} \hat{y}_{k/\beta}^2. \tag{210}$$

Here $\hat{y}_t$ is the FIR lowpass filter output, approximating the actual baseband signal values used in the estimate (209).

In this formulation, one lowpass filter is used whose passband is the entire baseband. One then samples the output signal at the Nyquist rate to get (in the Gaussian case) something close to statistically independent and identically distributed random variables. The number of these is approximately $N\beta$, the time bandwidth product.

5.4 Frequency Division Multiple Windows

The estimator of the previous section uses $N\beta$ output samples of a single lowpass filter of bandwidth β. The other extreme is to use an output sample from each of $N\beta$ narrowband filters, each having bandwidth $1/N$. This is the smallest bandwidth for which the Nyquist subsampling rate for the output signal would deliver at least one sample in a block of N consecutive values. Since we must limit ourselves to FIR filters of length N, we cannot achieve a perfect decomposition into narrow bands in this way. We shall see, however, that $N\beta$ is an upper bound on the number of length N FIR bandpass filters for which

(i) the pulse response sequences are orthogonal, and

(ii) the individual passbands are inside baseband.

For white Gaussian data, the first condition would make the output samples of any two bandpass filters uncorrelated and, therefore, statistically independent.

The simplest example of this approach is derived from the length N discrete Fourier transforms. We shall begin with this case and then attempt to generalize. To facilitate our discussion, suppose that

$$N\beta = m$$

is a positive odd integer. Construct V by letting its m rows consist (in any order) of normalized rows of the DFT matrix:

$$V^{(i)} = \frac{1}{N\sqrt{\beta}} \mathbf{1}^T D(e^{-j\frac{2\pi i}{N}}), \qquad |i| < \frac{m}{2}, \tag{211}$$

where $\mathbf{1}$ (see equation (149)) is the vector whose elements are all one. The normalization is chosen so that the estimator is unbiased in the white noise case or, equivalently,

$$\mathrm{Tr}[VV^*] = 1. \tag{212}$$

The row $V^{(i)}$ is the unit pulse response of an FIR bandpass filter with passband

$$\left|\theta - \frac{2\pi i}{N}\right| < \frac{\pi}{N} = \frac{\pi\beta}{m} \tag{213}$$

having normalized bandwidth $1/N$. The m rows of V are orthogonal and, in fact,

$$VV^* = \frac{1}{N\beta} I. \tag{214}$$

Therefore this example meets the two conditions we have specified, with a liberal interpretation of filter passband.

The spectrum estimator constructed with this choice of V is intimately related to the periodogram (to which it reduces when $m = 1$). Letting $\hat{S}_1$ be the periodogram of equation (147), our estimator becomes

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \frac{1}{m} \sum_{|i| < \frac{m}{2}} \hat{S}_1(e^{j[-2\pi i/N]}, \mathbf{x}). \tag{215}$$

To obtain a sample of $\hat{S}$, one takes the average of m samples of $\hat{S}_1$.

Toeplitz Quadratic Estimators. Generalizing the notion of averaging the periodogram leads to Toeplitz estimators. Let

$$H(e^{j\theta}) = \sum_{k=-\infty}^{\infty} h_k e^{-jk\theta} \tag{216}$$

be nonnegative real and concentrated in the vicinity of $\theta = 0$. Let

$$\hat{S}_1(e^{j\theta}, \mathbf{x}) = \frac{1}{N} \left| \mathbf{1}^T D(e^{-j\theta}) \mathbf{x} \right|^2 \tag{217}$$

be the periodogram, as in equation (144). Consider the averaged periodogram

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \frac{1}{\beta} \int_{-\pi}^{\pi} H(e^{j\phi}) \hat{S}_1(e^{j(\theta-\phi)}, \mathbf{x}) \frac{d\phi}{2\pi}. \tag{218}$$

This estimator is quadratic and has the form

$$\hat{S}(e^{j\theta}, \mathbf{x}) = \left|\left| V D(e^{-j\theta}) \mathbf{x} \right|\right|^2,$$

where

$$\begin{aligned} Q &= V^* V \\ &= \frac{1}{\beta N} \int_{-\pi}^{\pi} H(e^{j\phi}) D(e^{-j\phi}) \mathbf{1}\, \mathbf{1}^T D(e^{j\phi}) \frac{d\phi}{2\pi} \\ &= \frac{1}{\beta N} \int_{-\pi}^{\pi} H(e^{j\phi}) \Psi(e^{j\phi}) \Psi^*(e^{j\phi}) \frac{d\phi}{2\pi}. \end{aligned} \tag{219}$$

This matrix is nonnegative definite and *Toeplitz*:

$$Q_{kl} = \frac{1}{N} \int_{-\pi}^{\pi} H(e^{j\phi}) e^{-j\phi(k-l)} \frac{d\phi}{2\pi} = h_{l-k}. \tag{220}$$

If $H(e^{j\phi})$ is composed of m delta functions (and is therefore concentrated on m points, as it would be for the estimator of equation (215)), then Q has rank m. If, on the other hand, H is positive on an interval, then Q will be positive definite and have rank N. Thus, averaging the periodogram increases the estimator rank. (Recall that averaging the data, as in Section 5.1, does *not* increase rank.) All nonnegative definite Toeplitz matrices have the representation (219) or (220).

The spectrogram of Granander [19] is Toeplitz in our present sense. Consider once more the problems of estimating baseband power and then using this to estimate the power spectrum at $\theta = 0$ as in equation (202). The operator P in that equation is the approximate projection

$$P = P_T P_\Omega,$$

where P_Ω is the Toeplitz (lowpass filter) projection of equation (194) and P_T is the projection of rank N which zeroes out all sequence elements except for those in the range 0 to $N-1$. This can be written

$$P_T = F F^T,$$

where F is given by (204). It follows that

$$F^T P_T = F^T,$$

and therefore the quadratic form in equation (206) is

$$\begin{aligned} Q = V^T V &= \frac{1}{N\beta} F^T P_T P_\Omega F \\ &= \frac{1}{N\beta} F^T P_\Omega F. \end{aligned} \tag{221}$$

This matrix is a scaled $N \times N$ diagonal block of the Toeplitz prection operator P_Ω. It is also identical to the matrix of equation (219), provided $H(e^{j\theta})$ is the ideal baseband filter. And, finally, coming full circle, if R is the matrix of equation (17), then

$$Q = \frac{1}{N\beta} R. \tag{222}$$

Now Q is positive definite and has trace one. However, $N - m = N(1-\beta)$ of its eigenvalues are very close to zero, while the next are close to $1/m$. Thus, Q is close to a rank m matrix. Roughly speaking, this means that estimator error variance can be decreased by the factor $1/m$ only, even though the actual rank of Q is much larger.

Toeplitz quadratic forms are natural choices in view of the fact that every linear functional of the power spectrum is the expected value of a Toeplitz bilinear form in the infinite data sequence (Section 4.0). It has been argued here that a desirable quadratic form for estimating baseband power would have the properties

$$\begin{cases} Q = V^{\star}V \text{ is Toeplitz} \\ \operatorname{rank}(Q) = m = N\beta \\ VV^{\star} = \frac{1}{m}I \quad \text{(orthogonal filter responses).} \end{cases} \tag{223}$$

Our first example, equation (211), has these properties. In general, if $V^{\star}V$ is Toeplitz, then the estimator is a smoothed or averaged periodogram as in equation (218). If, in addition, it has rank m, then it is the average of m samples of the periodogram. Finally, it can be shown that, if $V^{\star}V$ is diagonal in addition to the other two properties, then the samples of the periodogram must be separated by integer multiples of $2\pi/N$. For the problem of estimating baseband power, these samples should be in baseband which has width $2\pi m/N$. Thus our example is essentially unique. Of course, multiplication on the left by a unitary matrix

$$V' = UV$$

will produce the same estimator. If Q satisfies the conditions of equations (223), then $P = mQ$ is a Toeplitz orthogonal projection. Thus

$$\begin{aligned} \hat{S}(e^{j\theta}, \mathbf{x}) &= \left|\left|VD(e^{-j\theta})\mathbf{x}\right|\right|^2 \\ &= \frac{1}{m}\left|\left|PD(e^{-j\theta})\mathbf{x}\right|\right|^2 \end{aligned} \tag{224}$$

is a projection-based estimator. Finally, every symmetric real Toeplitz matrix Q commutes with J. This is enough to make Toeplitz quadratic estimators J-symmetric in the sense of Section 4.3, since

$$Q = JQJ$$

implies

$$(VJ)^{\star}(VJ) = V^{\star}V,$$

which implies in turn that

$$VJ = UV$$

for some unitary matrix U.

Non-Toeplitz Forms. Now let us relax the requirement that $V^\star V$ be Toeplitz. We require the maximum number of orthogonal filters, whose passbands are in the baseband. Two orthogonal filters will give independent output samples in the Gaussian white data case. (Essentially, the power spectrum must be nearly constant on bands of normalized bandwidth β.) Therefore we can decrease error variance by the factor $1/m$, where m is the number of orthogonal filters we can squeeze into the baseband. What is required at this point is a precise statement of what this means.

Let

$$H(e^{j\theta}) = \sum_{k=0}^{N-1} h_k e^{-jk\theta} \tag{225}$$

be the frequency response function of an FIR filter. A minimum requirement, if this filter is to be a baseband filter, is that, for some choice of $\epsilon > 0$,

$$\int_{-\beta\pi}^{\beta\pi} \left|H(e^{j\theta})\right|^2 \frac{d\theta}{2\pi} \geq (1-\epsilon) \int_{-\pi}^{\pi} \left|H(e^{j\theta})\right|^2 \frac{d\theta}{2\pi}. \tag{226}$$

In other words, most of the filter energy should be in baseband. This is a quadratic inequality and can be written

$$\mathbf{h}^T R \mathbf{h} \geq (1-\epsilon)\mathbf{h}^T\mathbf{h}, \tag{227}$$

where

$$\mathbf{h} = [h_0 \; h_1 \cdots h_{N-1}]^T \tag{228}$$

and R is the Toeplitz baseband autocorrelation matrix of equation (17).

Each row of V corresponds to a filter response. The two conditions together lead to the following problem. Find the largest (meaning m) $m \times N$ matrix V for which

$$V\left[R - (1-\epsilon)I\right]V^\star \geq 0 \tag{229}$$

and

$$VV^\star = \frac{1}{m} I. \tag{230}$$

The eigenvalues of R are

$$\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_N^2 > 0,$$

and the eigenvalues of $R - (1 - \epsilon)I$ are

$$\lambda_1^2 - (1 - \epsilon), \ldots, \lambda_N^2 - (1 - \epsilon).$$

But we know the sum of these eigenvalues, using equation (17), to be

$$\mathrm{Tr}(R) = \sum_{i=1}^{N} \lambda_i^2 = N\beta. \tag{231}$$

Thus, for ϵ sufficiently small, at most $N\beta$ of the eigenvalues of R can be greater than $1 - \epsilon$. The inequality (217) cannot hold if fewer than m of the eigenvalues of $R - (1 - \epsilon)I$ are nonnegative. Therefore, $N\beta$ is an upper bound on the rank m for this problem (when ϵ is sufficiently small).

A stronger statement of this situation is found in the following.

Theorem: Let M be the $m \times m$ matrix on the left of inequality (229), and let M have eigenvalues

$$\mu_1^2 \geq \mu_2^2 \geq \cdots \geq \mu_m^2. \tag{232}$$

Then if equation (229) holds,

$$\mu_i^2 \leq \frac{\lambda_i^2 - (1 - \epsilon)}{m}, \tag{233}$$

for each i.

The proof of this theorem involves a generalized Rayleigh quotient inequality argument. Of course, M is nonnegative definite if and only if all m of its eigenvalues are nonnegative. Equality in (233) can be obtained by letting the columns of V^* be eigenvectors of R, in particular those having the m largest eigenvalues. This choice produces the Thomson spectral estimator, and J is symmetric. (This theorem addresses the problem in Section 2.5.)

In conclusion, we have argued that, for a specified resolution β, there is a limit to the reduction in error variance possible with multiple window quadratic estimators. The rank m should be approximately the time bandwidth product $N\beta$.

6.0 CONCLUSIONS

The basic problem in classical spectrum analysis is to project a time series onto a timelimited and bandlimited subspace where power can be estimated. Such a subspace can only be approximated, so the problem can be rephrased as one of constructing approximating subspaces and projections onto them. The first obvious approach is to build a frequency selective FIR filter, and the natural extension of this approach is to build a frequency selective linear transformation. In Section 2.0 of this chapter we have followed this approach to its logical conclusion and found Slepian sequences as the appropriate sequences for building a subspace that is timelimited and approximately bandlimited. In Section 3.0 we have rephrased the problem of spectrum analysis as one of estimating parameters in a structured covariance matrix. Maximum likelihood estimates of these parameters produce spectrum estimators which are essentially equivalent to the multiwindow spectrum estimators of Thomson and to rank reduced versions of Daniell's frequency averaged periodogram. The maximum likelihood spectrum estimator is a quadratic form in the data that is formed by complex demodulating the time series, projecting it onto a low-rank subspace, and computing its power in that subspace. The mean-squared error of the estimator decreases inversely with the rank of the quadratic form that is constructed in this way. In Section 4.0 we show that every quadratic estimator of the power spectrum that is required to be nonnegative and modulation invariant must be a quadratic form in a complex demodulated time series. This fundamental representation theorem characterizes a class of admissible spectrum estimators. The maximum likelihood estimators are members of this class, but there are others. They include all of the windowed and frequency averaged periodograms and others discussed in Section 5.0.

From the point of view of this chapter, there is nothing very fundamental about uniform sampling, nor is there anything very fundamental about a scalar index parameter for the data. This means that our results extend to the frequency-wavenumber analysis of nonuniformly sampled space-time series. These extensions, developed somewhat in [10] and [11], form the basis for a research program in the spectrum analysis of multiparameter data sequences.

Acknowledgements

The authors acknowledge stimulating discussions with B. D. Van Veen and Li Du. This work supported by the Office of Naval Research under contract No. N00014-89-J-1070, by the Army Research Office under contract No. DAAL03-88-K-0058, and by the National Science Foundation under contract No. MIP-8717714.

References

[1] A. Schuster, "On Lunar and Solar Periodicities of Earthquakes," *Proc R Soc* **61**, pp. 455–465 (1897).

[2] A. Einstein, "Method for the Determination of the Statistical Values of Observations Concerning Quantities Subject to Irregular Fluctuations," *Arch Sci Phys et Natur* **37**:4, pp. 254–256 (1914) See W. A. Garder, "Introduction to Einstein's Contribution to Time Series Analysis," *IEEE ASSP Mag* **4**:4, pp. 4–5 (1987).

[3] N. Wiener, "Generalized Harmonic Analysis," *Acta Math* **55**, pp. 117–258 (1930).

[4] R. B. Blackman and J. W. Tukey, "The Measurement of Power Specta from the Point of View of Communications Engineering," *BSTJ* **33**, pp. 185–282 and pp. 485–569 (1958).

[5] P. J. Daniell, "Discussion on the Symposium of Autocorrelation in Time Series, *J R Stat Soc* (Suppl.) **8**, pp. 88–90 (1946).

[6] P. D. Welch, "The Use of the Fast Fourier Transform for the Estimation of Power Spectra," *IEEE Trans Audio Electroacoust* **15**, pp. 70–73 (1970).

[7] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty–V: The Discrete Case," *BSTJ* **57**, pp. 1371–1430 (May 1978).

[8] D. J. Thomson, "Spectrum Estimation and Harmonic Analysis," *Proc IEEE* **70**:9, pp. 1055–1096 (1982).

[9] H. Clergeot, "Choix Entre Différentes Méthodes Quadratiques d'Estimation Spectrale," *Ann Télécommun* **39**:3,4, pp. 113–128 (1984).

[10] T. P. Bronez, "Spectral Estimation of Irregularly Sampled Multidimensional Processes by Generalized Prolate Spheroidal Sequences," *IEEE Trans ASSP* **36**, pp. 1862–1873 (1988).

[11] L. L. Scharf and Li Du, "Components of Variance Interpretation of Spectrum Estimation," *Proc Nat Radio Sci Mtg*, Boulder, Colorado (January 1988).

[12] J. P. Burg, "Maximum Entropy Spectral Analysis," *Proc 37th Mtg Soc of Explor Geophysicists* (1967).

[13] C. J. Gueguen and L. L. Scharf, "Exact Maximum Likelihood Identification of ARMA Models: A Signal processing Perspective," *Proc EUSIPCO-80*, Lausanne, pp. 759–769 (September 1980).

[14] E. Parzen, "Some Recent Advances in Time Series Modelling," *IEEE Trans AC* **AC-19**, pp. 723–730 (December 1974).

[15] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of Structured Covariance Matrices," *Proc IEEE* **70**, pp. 963–974 (September 1982).

[16] L. L. Scharf and P. J. Tourtier, "Maximum Likelihood Identification of Structured Correlation Matrices for Spectrum Analysis," *Proc 20th Asilomar Conf on Signals, Systems, and Computers* (November 1986).

[17] P. J. Tourtier and L. L. Scharf, "Maximum Likelihood Identification of Correlation Matrices for Estimation of Power Spectra at Arbitrary Resolutions," *Proc ICASSP 87*, pp. 2066–2069 (April 1987).

[18] B. D. Van Veen and L. L. Scharf, "Spectrum Analysis Using Low Rank Models and Structured Covariance Matrices," *Proc EUSIPCO-88*, Grenoble (1988).

[19] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series* (New York: John Wiley & Sons, Inc., 1917).

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS None
2a. SECURITY CLASSIFICATION AUTHORITY DISCASS	3. DISTRIBUTION/AVAILABILITY OF REPORT Unrestricted
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CU/ONR/2	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a. NAME OF PERFORMING ORGANIZATION Dept. Elect & Computer Eng. University of Colorado 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research Mathematics Division
6c. ADDRESS (City, State and ZIP Code) Campus Box 425 Boulder, CO 80309-0425	7b. ADDRESS (City, State and ZIP Code) 800 North Quincy Street Arlington, VA 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) Quadratic Estimators of the Power Spectrum (U)	

12. PERSONAL AUTHOR(S)
C. T. Mullis and Louis L. Scharf

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ____ TO ____	April 1989	78

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Among nonparametric estimators of the power spectrum, quadratic estimators are the only ones that are dimensionally correct. In this report we motivate interest in quadratic estimators by setting up idealized experiments for spectrum analysis and deriving maximum likelihood estimates for narrowband power. These idealized experiments show that quadratic functions of the experimental data are sufficient statistics for estimating power. The maximum likelihood estimates show that low-rank projection operators play a fundamental role in the theory of spectrum analysis. The projection operator plays the same role as a band-pass filter in a conventional swept frequency spectrum analyzer. The mean-squared error of a maximum likelihood estimator of power is inversely proportional to the rank of the estimator.

With our maximum likelihood results in hand, we turn to a systematic study of quadratic estimators of the power spectrum. We prove a fundamental representation theorem for any quadratic estimator of the power spectrum that is required to be positive and modulation-

(see reverse)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☒	21. ABSTRACT SECURITY CLASSIFICATION U

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Louis L. Scharf	(303) 492-8283	

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DISTRIBUTION:

Dr. Neil L. Gerr
Statistics Branch
Mathematics Division
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Max Irving
Resident Representative
Office of Naval Research
University of New Mexico
Room 223, Bandolier Hall West
Albuquerque, NM 87131

Director, Naval Research Laboratory
Attn: Code 2627
Washington, DC 20375

Defense Technical Information Center
Building 5
Cameron Station
Alexandria, VA 22314

Martha Campbell
Government Publications
University of Colorado
Campus Box 184
Boulder, CO 80309-0184

Office of Contracts and Grants
University of Colorado
Campus Box 19
Boulder, CO 80309-0019